Submitted to the Annals of Statistics
arXiv: math.PR/0000000

BAYESIAN MISE BASED CONVERGENCE RATES OF MIXTURE MODELS BASED ON THE POLYA URN MODEL: ASYMPTOTIC COMPARISONS AND CHOICE OF PRIOR PARAMETERS

BY SABYASACHI MUKHOPADHYAY[*,‡]
AND SOURABH BHATTACHARYA[†,‡]

Indian Statistical Institute[‡]

We study the asymptotic properties of Bayesian density estimators constructed using normal mixtures of Dirichlet process priors where the random probability measure has been integrated out, so that the mixture is induced by the Polya urn scheme (Blackwell and McQueen (1973)). Thus, the density estimators we consider consist of at most finite number of mixture components for a given sample size, even though the number of mixture components increase with increasing sample size. This is in contrast with the existing works on Bayesian density estimation (for example, Ghosal and van der Vaart (2007)) where the random measure is retained, while parameters arising from the random measure are assumed to be integrated out (in principle). Within our marginalized mixture framework, we consider two separate density estimators; that of Escobar and West (1995) and that introduced by Bhattacharya (2008). The latter mixture model specifies a bound to the number of mixture components, preventing it from growing arbitrarily large with the sample size. We study the posterior rates of convergence of the mean integrated squared error ($MISE$) for both kinds of mixtures and show that the $MISE$ corresponding to Bhattacharya (2008) converges to zero at a much faster rate compared to that of Escobar and West (1995). We also show that with proper, but plausible, choices of the free parameters of our $MISE$ bounds the rate of convergence can be made smaller than the best rate of Ghosal and van der Vaart (2007) given by $n^{-2/5}(\log n)^{4/5}$ and in fact, can be made smaller that the optimal frequentist rate $n^{-2/5}$. Apart from these we study and compare the $MISE$ convergence rates of the two models in the case of the "large p small n" problem. Furthermore, we show that while the model of Escobar and West (1995) can converge to a wrong model under certain conditions, much stonger conditions are necessary for Bhattacharya (2008) to converge to a wrong model. Finally, we consider a modified version of Bhattacharya (2008) but demonstrate that all the results remain same under the modified version.

[*]Sabyasachi Mukhopadhyay is a PhD student in Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute

[†]Sourabh Bhattacharya is an Assistant Professor in Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute

AMS 2000 subject classifications: Primary 62F15; secondary 62G07

Keywords and phrases: Bayesian Asymptotics, Dirichlet Process, Mean Integrated Squared Error, Mixture Analysis, Polya Urn

1. Introduction. In recent years, the use of nonparametric prior in the context of Bayesian density estimation arising out of mixtures has received wide attention thanks to their flexibility and advances in computational methods. The study of nonparametric priors in the context of Bayesian density estimators has been initiated by Ferguson (1983) and Lo (1984) who derived the associated posterior and predictive distributions. The set-up of for nonparametric Bayesian density estimation can be represented in the following hierarchical form: for $i = 1, \ldots, n$, $Y_i \sim K(\cdot \mid \theta_i)$ independently; $\theta_1, \ldots, \theta_n \overset{iid}{\sim} F$ and $F \sim \Pi$, where F is a random probability measure and Π is some appropriate nonparametric prior distribution on the set of probability measures. An important choice of Π is of course the Dirichlet process prior, for which the set-up of Lo (1984) boils down to the Escobar and West (1995) (henceforth EW) model.

Though very well known, the EW model has several draw backs in terms of computational efficiency (see Mukhopadhyay, Roy and Bhattacharya (2012), Mukhopadhyay, Bhattacharya and Dihidar (2011)) which manifest themselves particularly when applied to massive data. Bhattacharya (2008) (henceforth SB) proposed a new model which is shown to bypass the problems of the EW model (see Mukhopadhyay, Bhattacharya and Dihidar (2011) and Mukhopadhyay, Roy and Bhattacharya (2012)). The essence of the SB model lies in the assumption that data points are independently and identically distributed (iid) as an M-component mixture model, where the parameters of the mixture components, which we denote by $\theta_1, \ldots, \theta_M$, are samples from a Dirichlet process. Thus, the total number of distinct components of the mixture is bounded above by M. If M is chosen to be much less than n, the number of data points, then this idea entails great computational efficiency compared to the EW model, particularly in the case of massive data. Moreover, if $M = n$, and Y_i is associated with θ_i for every i, then the SB model reduces to the EW model, showing that the EW model is a special case of the SB model. In this paper, we will assume M to be increasing with n; in fact, our subsequent asymptotic calculations show that M increasing at a rate slower that $\sqrt{n}$, is adequate. To reflect the dependence of M on n henceforth we shall write M_n. Thus, the dimensionality of both the EW and the SB model grows with n, although in the latter case it grows at a slower rate.

However, the problem of comparison between these two models with respect to asymptotic posterior convergence rates of the random density estimators has not been addressed. Although Ghosal and van der Vaart (2007) addressed the consistency issues of the EW model and obtained the posterior convergence rates, these are obtained under the assumption that randomness of the kernel mixture is induced by the random measure F while the parameters θ_i are integrated out (in principle). Consequently, hitherto all the asymptotic calculations on random densities are done with respect to the random measure F.

Our interest, on the other hand, lies in posterior consistency and convergence rates of random densities under the set-up where randomness of the kernel mixture is induced via the parameters, assuming that the random measure F has been integrated out. Indeed, in almost all practical settings, F, which is typically assumed to have the Dirichlet process prior, is not of interest and is integrated out; the kernel mixture is rendered random only through the unknown parameters, which, in the case of Dirichlet process prior, follow the Polya urn scheme. Both EW and SB exploit the properties of the Polya urn distribution to construct their models and associated methodologies.

Assuming the aforementioned set-up where F is integrated out, in this paper we investigate posterior consistency and convergence rates of EW and SB models, using $MISE$ as the divergence measure between the kernel mixture and the true density which is assumed to generate the data. In particular, we show that the model of SB converges much faster than that of EW, in terms of $MISE$ based on the posteriors.

Optimization of the $MISE$ convergence rates helps obtain asymptotical optimal choices of several important prior parameters driving the models. This is important since in applications the prior parameters are almost always chosen by *ad hoc* means.

We also study the rates of convergence of the two competing models in the "large p small n" set-up, and show how the priors must be adjusted to guarantee consistency in this case. As before, in this set-up as well the SB model beats the EW model in terms of faster rate of $MISE$-based posterior convergence.

There is also an important question regarding the conditions leading to convergence of the mixtures to the wrong models (that is, models that did not generate the data). We show that the model of EW can converge to a wrong model under relatively weak conditions, whereas much stronger conditions must be enforced to get the SB model to converge to the wrong model.

Furthermore, we consider a modified version of SB's model; however, as we demonstrate, all the results remain intact under this version.

The rest of the work is organized as follows. In Sections 2 and 3 we provide details of the explicit forms of the EW-based and the SB-based density estimators and provide discussions on the assumptions used in our subsequent asymptotic calculations. The assumptions regarding the true, data-generating distribution are provided in Section 4. Section 5 provides details of our Bayesian $MISE$-based divergence measure used for our asymptotic calculations. Sections 6 and 7 provide results showing convergence of the posterior expectations of the EW-based and the SB-based density estimators, respectively, to the same true distribution, also providing the rates of convergence. In Sections 8 and 9 we compute the *a posteriori* Bayesian $MISE$-based rates of convergence of the EW and the SB model.

In Section 10 the posterior rates of the two models are compared, and the "large p small n problem" is investigated in Section 12. In Section 13, the conditions, under which the models may converge to wrong distributions, is investigated. The convergence rate of a modified version of SB model is calculated in Section 14. Proofs of most of the results are provided in the supplement Mukhopadhyay and Bhattacharya (2012b), whose sections have the prefix "S-" when referred to in this paper. Additionally, in Section S-9 of the supplement, we study $MISE$ rates of convergence with respect to the prior– since the number of parameters in these two models depends upon the sample size (in the EW model the number of parameters is the same as the sample size, and in the SB model the number of parameters is M, which we will assume to increase with the sample size), it is possible, by increasing the number of parameters, to obtain prior rates of convergence. We compare the $MISE$-based prior and the posterior convergence rates in Section S-10 of the supplement.

2. The EW model and the associated assumptions. We assume the following version of the EW model: $[Y_i \mid \theta_i, \sigma] \sim N(\theta_i, \sigma^2)$ (normal distribution with mean θ_i and variance σ^2), $\theta_i \overset{iid}{\sim} F$ for all i, $F \sim D(\alpha G_0)$, where we assume $G_0(\theta) \equiv N(\mu_0, \sigma_0^2)$, where μ_0 and σ_0 are known. We let the parameter α increase with n, the sample size. Further we assume a sequence of priors on σ as $\sigma/\sigma_n \sim G$, where σ_n is a sequence of constants such that $0 < \sigma_n < K^*$, where K^* is finite; $\sigma_n \to 0$, and G is fixed. Let $\sigma \sim G_n$, where $G_n(s) = G(s/\sigma_n)$. This assumption regarding the prior of σ is very similar to that of Ghosal and van der Vaart (2007). Following Ghosal and van der Vaart (2007) we also assume that $P(\sigma > \sigma_n) = O(\epsilon_n)$, where $\epsilon_n \to 0$. As we make precise later, we let the choice of ϵ_n depend upon the other prior parameters. We let $\Theta_n = (\theta_1, \ldots, \theta_n)'$. It is important to note that, for a particular value of n, we have a particular form of prior of σ, given by G_n. Thus, unlike a single sequence of random variables we have a *double array* of random variables, as illustrated below:

$$
\begin{array}{llll}
Y_{11} & Y_{12} \ldots Y_{1k_1} & & \\
Y_{21} & Y_{22} & \ldots & Y_{2k_2} \\
 & \cdot & & \\
 & \cdot & & \\
Y_{n1} & Y_{n2} & \ldots & Y_{nk_n} \\
 & \cdot & & \\
 & \cdot & &
\end{array}
$$

For each n, there are k_n random variables $\{Y_{ni}, i = 1, \ldots, k_n\}$. It is assumed that $k_n \to \infty$ as $n \to \infty$. In the array Y_{ij}'s are assumed to be independent among

each other. In the case of the EW model, $k_n = n$, a situation in which the double array of random variables is referred to as the *triangular array* of random variables (see, for example, Serfling (1980)). In the EW model, for each n, $Y_{ni} \sim N(\theta_i, \sigma^2)$ and the prior specifications on θ_i and σ are the same as stated in the beginning of this section. For notational simplicity we drop the suffix "n" in $\{Y_{ni}; i = 1, \ldots, n\}$ and simply denote it by $\{Y_1, \ldots, Y_n\}$.

We study the posterior-based asymptotic properties of prior predictive densities of the following form:

$$\hat{f}_{EW}(y \mid \Theta_n, \sigma) = \frac{\alpha}{\alpha+n} A_n + \frac{1}{\alpha+n} \sum_{i=1}^{n} \frac{1}{(\sigma+k)} \phi\left(\frac{y-\theta_i}{\sigma+k}\right), \tag{1}$$

where $\phi(\cdot)$ is the standard normal density, and $A_n = \int_\theta \frac{1}{(\sigma+k)} \phi\left(\frac{y-\theta}{\sigma+k}\right) dG_0(\theta)$. Addition of the positive constant k to the standard deviation is essentially a device to prevent the variance from getting arbitrarily close to zero. It is important to note the difference between the model assumption for the data and the density estimator of our interest given by (1); even though the latter adds k to σ, the former does not consider addition of any positive constant to σ. In spite of slightly inflating the variance in (1) the form of the true distribution (4) to which (1) converges *a posteriori* is not severely restricted.

3. SB model and the associated assumptions. As in case of EW model, here we consider the following density estimator:

$$\hat{f}_{SB}(y \mid \Theta_{M_n}, \sigma) = \frac{1}{M_n} \sum_{i=1}^{M_n} \frac{1}{(\sigma+k)} \phi\left(\frac{y-\theta_i}{\sigma+k}\right), \tag{2}$$

where M_n is the maximum number of distinct components the mixture model can have and $\Theta_{M_n} = (\theta_1, \theta_2, \ldots, \theta_{M_n})'$. Define $Z_i = j$ if Y_i comes from the j-th component of the mixture model. Denote z as the realized vector of Z. We make the same assumptions regarding α and σ as in EW. Additionally, we let M_n increase with n. Same as in case of EW model, $\boldsymbol{Y}_n$ for SB model also forms a triangular array.

To perform our asymptotic calculations with respect to the SB model, we need to shed light on an issue associated with the frequentist estimate of σ_T^2, the variance of the true density f_0 generating the data. The assumptions on the true distribution f_0 are provided in Section 4.

Let $n_j = \#\{t : z_t = j\}$, $n = \sum_{j=1}^{M_n} n_j$, $\hat{\sigma}^2_{T,n} = \frac{\sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2}{n} = \frac{\sum_{j=1}^{M_n} n_j s^2_{j,n}}{n}$, where $s^2_{j,n} = \frac{\sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2}{n_j}$. Now, defining $\bar{Y} = \frac{\sum_{j=1}^{M_n} n_j \bar{Y}_j}{n}$ we note

that $\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2$ can be expressed, for any allocation vector $z=(z_1,\ldots,z_n)'$, as

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2 &= \frac{1}{n}\sum_{j=1}^{M_n}\sum_{t:z_t=j}(Y_t-\bar{Y})^2\\ &= \frac{1}{n}\sum_{j=1}^{M_n}n_j\left(\bar{Y}_j-\bar{Y}\right)^2+\hat{\sigma}^2_{T,n}.\end{aligned}$$

Since $\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2\rightarrow\sigma^2_T$ a.s., it would follow from the above representation that $\hat{\sigma}^2_{T,n}\rightarrow\sigma^2_T$ a.s. if it can be shown that $\frac{1}{n}\sum_{j=1}^{M_n}n_j\left(\bar{Y}_j-\bar{Y}\right)^2\rightarrow 0$ a.s. The following lemma guarantees that it is indeed the case.

LEMMA 3.1. *Under the data generating true density f_0, $\frac{1}{n}\sum_{j=1}^{M_n}n_j\left(\bar{Y}_j-\bar{Y}\right)^2\rightarrow 0$, a.s. if $1<M_n\prec O(n)$ (for any two sequences $a^{(1)}_n$ and $a^{(2)}_n$ we say $a^{(1)}_n\prec a^{(2)}_n$ if $\frac{a^{(1)}_n}{a^{(2)}_n}\rightarrow 0$).*

PROOF. See Section S-1.1 of the supplement.

□

From lemma 3.1 and the fact that $\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2\rightarrow\sigma^2_T$ a.s. we can conclude $\hat{\sigma}^2_{T,n}\rightarrow\sigma^2_T$, a.s. So, as $n\rightarrow\infty$, $n\hat{\sigma}^2_{T,n}\sim n\sigma^2_T$ a.s., (for any two sequences $\{a^{(1)}_n\}$ and $\{a^{(2)}_n\}$, $a^{(1)}_n\sim a^{(2)}_n$ denotes $\lim_{n\rightarrow\infty}\frac{a^{(1)}_n}{a^{(2)}_n}=1$), implying that as $n\rightarrow\infty$, $\sum_{j=1}^{M_n}\sum_{t:z_t=j}(Y_t-\bar{Y}_j)^2$ becomes independent of z. We begin by writing $n\hat{\sigma}^2_{T,n}\sim C_n$, where $0<\frac{C_n}{n}<\aleph$ (for some sufficiently large constant $\aleph$) is a bounded sequence independent of z and has the same limiting behaviour as $\hat{\sigma}^2_{T,n}$. We will perform our calculations when for each n, $|Y_i|<a; i=1,\ldots,n$, for some sufficiently large constant $a>0$. In this case $0<\hat{\sigma}^2_{T,n}<4a^2$. Thus, we may choose $\aleph=4a^2$.

To prove many of our results on the SB model, we will assume suitable conditions on C_n. The conditions on C_n that we assume are reasonable, and are consistent with the above discussion. In keeping with the above discussion, the proof of Lemma 7.4 requires us to assume that for large n, C_n/n is greater than or equal to the order of $\sigma^2_n\log\left(\frac{1}{\sigma_n}\right)$, unless ϵ_n tends to zero at too fast a rate compared to σ^2_n. In the latter situation σ is much harder to estimate since in that case the sample size n will be smaller, σ_n larger, and σ can be anything between 0 and σ_n, which is to be estimated with a relatively small sample. Hence, although in the former case we expect C_n/n get close to zero for large n, in the latter case C_n/n may remain bounded away from zero.

4. Assumptions regarding the true distribution. In this paper we assume that the true, data generating distribution is of the following form:

$$f_0(y) = \int_{-a-c}^{a+c} \frac{1}{k}\phi\left(\frac{y-\theta}{k}\right) dF_0(\theta), \tag{3}$$

where k is some known positive constant, and F_0 is an unknown distribution compactly supported on $[-a-c, a+c]$, for some constants $a > 0$ and $c > 0$.

Note that, using the mean value theorem for integrals, also known as the general mean value theorem (GMVT) we can re-write $f_0(y)$ as

$$f_0(y) = \frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right), \tag{4}$$

where $\theta^*(y) \in (-a-c, a+c)$ may depend upon y.

The results in the following two sections show that the posterior expectations of the Bayesian density estimators corresponding to the EW and the SB models, given by (1) and (2) converge to the form given by (4).

5. Bayesian MISE based posterior convergence. Assuming that $\hat{f}_n$ is an estimate of the true density f_0 based on the observed data $\boldsymbol{Y}_n = (Y_1, \ldots, Y_n)'$, the MISE of $\hat{f}_n$ is given by

$$MISE = \int_y E(\hat{f}_n(y) - f_0(y))^2 dy, \tag{5}$$

where the expectation is with respect to the data $\boldsymbol{Y}_n$. In our Bayesian context, we assume that the density estimate of $f_0(y)$ is $\hat{f}(y|\Theta)$, where Θ is an unknown set of parameters. We consider the following analogue of the above classical definition:

$$\begin{aligned} MISE_1 &= \int_y E(\hat{f}(y|\Theta) - f_0(y))^2 dy \\ &= \int_y \int_\Theta \{\hat{f}(y|\Theta) - f_0(y)\}^2 [\Theta \mid \boldsymbol{Y}_n] d\Theta dy, \end{aligned} \tag{6}$$

that is, the expectation is with respect to the posterior of Θ, generically denoted by $[\Theta|\boldsymbol{Y}_n]$. We further modify definition (6) by considering a weighted version, given by

$$\begin{aligned} MISE_2 &= \int_y E(\hat{f}(y|\Theta) - f_0(y))^2 f_0(y) dy \\ &= \int_y \int_\Theta \{\hat{f}(y|\Theta) - f_0(y)\}^2 [\Theta \mid \boldsymbol{Y}_n] f_0(y) d\Theta dy, \end{aligned} \tag{7}$$

where $f_0(y)$ is the weight associated with $\{\hat{f}(y|\Theta) - f_0(y)\}^2$. Thus, $f_0(y)$ downweights those squared error terms $\{\hat{f}(y|\Theta) - f_0(y)\}^2$ which correspond to extreme values of y. Such weighting strategies that use the true distribution as weight, are not uncommon in the statistical literature. The well-known Cramér-von Mises test statistic (see, for example, Serfling (1980)) is a case in point. For notational simplicty we refer to $MISE_2$ simply as $MISE$.

$MISE$ of the form (7) can be expressed conveniently as

$$\begin{aligned} MISE &= \int_y Var\left(\hat{f}\,(y \mid \Theta)\,|\boldsymbol{Y}_n\right) f_0(y)dy \\ &\qquad + \int_y \left\{Bias(\hat{f}(y \mid \Theta)|\boldsymbol{Y}_n)\right\}^2 f_0(y)dy, \end{aligned} \tag{8}$$

where $Var\left(\hat{f}\,(y \mid \Theta)\,|\boldsymbol{Y}_n\right)$ denotes the variance of $\hat{f}\,(y \mid \Theta)$ with respect to the posterior $[\Theta \mid \boldsymbol{Y}_n]$ and

$$Bias(\hat{f}(y \mid \Theta)|\boldsymbol{Y}_n) = \left|E\left(\hat{f}(y \mid \Theta)\middle|\boldsymbol{Y}_n\right) - f_0(y)\right|, \tag{9}$$

$E\left(\hat{f}(y \mid \Theta)\middle|\boldsymbol{Y}_n\right)$ denoting the expectation of $\hat{f}(y \mid \Theta)$ with respect to $[\Theta \mid \boldsymbol{Y}_n]$.

For the EW and the SB models we will denote the respective $MISE$'s as $MISE(EW)$ and $MISE(SB)$, respectively. Let $\boldsymbol{S}_n = \{\boldsymbol{Y}_n : \max_{1\leq i\leq n}|Y_i| < a\}$, and let $\mathbb{I}_{\boldsymbol{S}_n}$ denote the indicator function of the set $\boldsymbol{S}_n$. Also, let E_0^n denote the expectation with respect to f_0, the true distribution of $\boldsymbol{Y}_n$. We will compute and compare the rates of convergence to 0 of $E_0^n\left[MISE(EW)\mathbb{I}_{\boldsymbol{S}_n}\right]$ and $E_0^n\left[MISE(SB)\mathbb{I}_{\boldsymbol{S}_n}\right]$ when the true density f_0 is estimated using the EW model and the SB model, but with the same set of data for any given sample size. In particular, we show that $E_0^n\left[MISE(SB)\mathbb{I}_{\boldsymbol{S}_n}\right]/E_0^n\left[MISE(EW)\mathbb{I}_{\boldsymbol{S}_n}\right] \rightarrow 0$.

Before proceeding to the $MISE$ calculations, we first investigate the asymptotic forms of the posterior expectations of the EW-based and the SB-based density estimators given by (1) and (2), respectively. This we do in the next two sections. These results, apart from being interesting in their own rights and showing explicitly the form of the true distribution (the asymptotic form of posterior expected density estimators), actually provide the orders of the bias terms of the corresponding $MISE$ calculations, recalling from (8) that $MISE$ can be broken up into a variance part and a bias part.

Henceforth we will denote $\int_{-a-c}^{a+c} dG_0(x)$ by H_0. In proving most of our results the GMVT will play a very important role.

6. Convergence of the posterior expectation of the EW density estimator to the true distribution.

THEOREM 6.1. *Under the assumptions stated in Sections 2, 4, and 5, for any* $\boldsymbol{Y}_n \in \boldsymbol{S}_n$,

$$\sup_{-\infty<y<\infty}\left|E\left(\hat{f}_{EW}(y \mid \Theta_n,\sigma)\middle|\boldsymbol{Y}_n\right)-f_0(y)\right| = O\left(\frac{\alpha}{\alpha+n}+\frac{n}{\alpha+n}(B_n+\epsilon_n^*+\sigma_n)\right), \tag{10}$$

where

$$B_n = \frac{\alpha+n}{\alpha}e^{-\frac{c^2}{4\sigma_n^2}}, \tag{11}$$

and

$$\epsilon_n^* = \frac{\epsilon_n}{1-\epsilon_n}e^{\left(\frac{n(a+c_1)^2}{2{b_n}^2}\right)}\frac{(\alpha+n)^n}{(\alpha)^n}\frac{1}{H_0^n}. \tag{12}$$

In (12) $\{b_n\}$ *is a sequence of positive numbers such that* $0<b_n<\sigma_n$ *for all* n, $\alpha = O(n^{\omega})$, $0<\omega<1$, $\mu^*(y)\in(-c_1,c_1)$ *for each* y, *and* $f_0(y)=\frac{1}{k}\phi\left(\frac{y-\mu^*(y)}{k}\right)$ *is a well-defined density. The constant involved in the order (10) is independent of* $\boldsymbol{Y}_n$.

PROOF. See Section S-2.3 of the supplement. The proof depends upon several lemmas, stated below. □

Remark 1: We will choose ϵ_n such that $\epsilon_n^*\to 0$; in other words, we choose ϵ_n such that $\frac{\epsilon_n}{1-\epsilon_n}\prec\left[e^{\frac{n(a+c_1)^2}{2{b_n}^2}}\frac{(\alpha+n)^n}{(\alpha)^nH_0^n}\right]^{-1}$.

Remark 2: Note that, since the right hand side of (10) does not depend upon $\boldsymbol{Y}_n\in\boldsymbol{S}_n$, it follows that

$$E_0^n\left[\sup_{-\infty<y<\infty}\left|E\left(\hat{f}_{EW}(y \mid \Theta_n,\sigma)\middle|\boldsymbol{Y}_n\right)-f_0(y)\right|\mathbb{I}_{\boldsymbol{S}_n}\right] = O\left(\frac{\alpha}{\alpha+n}+\frac{n}{\alpha+n}(B_n+\epsilon_n^*+\sigma_n)\right).$$

Remark 3: The proof of Theorem 6.1 shows that for each y, $\mu^*(y)$ corresponds to $\sigma=0$ (the limit of the sequence σ_n), and so $\mu^*(y)$ is non-random.

The terms ϵ_n^* and B_n arise as the orders of the posterior probabilities $P(\sigma > \sigma_n|\boldsymbol{Y}_n)$ and $P\left(\theta_i \in [-a-c, a+c]^c, \sigma \leq \sigma_n|\boldsymbol{Y}_n\right)$, respectively. The first term in the order (19) of Theorem 8.1 is contributed by the order of the term $\frac{\alpha}{\alpha+n}A_n$, where A_n is already defined in connection with (1). These results, used for proving Theorem 6.1, which also play important roles in proving our main Theorem 8.1 on $MISE$ related to the EW model, our main theorem on $MISE$ in this section, are made precise in the following lemmas.

LEMMA 6.2. *Let $\{b_n\}$ and $\{\sigma_n\}$ be sequences of positive numbers such that $\sigma_n \to 0$, $0 < b_n < \sigma_n$ for all n, and $P(\sigma > \sigma_n) = O(\epsilon_n)$, for some sequence of positive constants $\epsilon_n > 0$. If $|Y_i| < a; i = 1, \ldots, n$, then*

$$P(\sigma > \sigma_n|\boldsymbol{Y}_n) = O\left(\epsilon_n^*\right) \tag{13}$$

PROOF. See Section S-2.1 of the supplement. □

LEMMA 6.3. *Under the same assumptions as Lemma 6.2 and $0 < \sigma_n < a$, the following holds:*

$$P\left(\theta_i \in [-a-c, a+c]^c, \sigma \leq \sigma_n|\boldsymbol{Y}_n\right) \leq \frac{C_0}{e^{-1/2}\delta}B_n, \tag{14}$$

where $C_0 = \sup_\sigma\{\sigma^{-1}\exp(-c^2/4\sigma^2)\}$, and δ is the lower bound of the density of G_0 on $[-a-c, a+c]$.

PROOF. This proof is similar to that of Lemma 11 of Ghosal and van der Vaart (2007). □

LEMMA 6.4. $\frac{\alpha}{\alpha+n}A_n = O\left(\frac{\alpha}{\alpha+n}\right)$, *for $\alpha = O(n^\omega)$, $0 < \omega < 1$.*

PROOF. See Section S-2.2 of the supplement. □

7. Convergence of the posterior expectation of the SB density estimator to the true distribution.

THEOREM 7.1. *Under the assumptions stated in Sections 3, 4, and 5, the following holds for any $\boldsymbol{Y}_n \in \boldsymbol{S}_n$:*

$$\begin{aligned}
&\sup_{-\infty<y<\infty}\left|E\left(\hat{f}_{SB}(y \mid \Theta_{M_n}, \sigma)\middle|\boldsymbol{Y}_n\right) - \frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)\right| \\
&= O\left(M_nB_{M_n} + \left(1-\frac{1}{M_n}\right)^n\left(\frac{\alpha+M_n}{\alpha}\right) + \epsilon_{M_n}^* + \sigma_n\right),
\end{aligned} \tag{15}$$

where $\epsilon^*_{M_n} = \frac{\epsilon_n}{1-\epsilon_n}\exp\left(\frac{n(a+c_1)^2}{2(b_n)^2}\right)\frac{(\alpha+M_n)^{M_n}}{\alpha^{M_n}H_0^{M_n}}$, $B_{M_n} = \frac{(\alpha+M_n)}{\alpha}e^{\left(-\frac{c^2}{4\sigma_n^2}\right)}$, *and* b_n *is as defined in Remark 2 associated with Theorem 6.1. Also, for every* y, $\theta^*(y) \in (-a-c, a+c)$, *and* E_0^n *denotes the expectation with respect to the true distibution of* $\boldsymbol{Y}_n$, *given by* $f_0(y) = \frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)$. *The constant involved in the above order is independent of* $\boldsymbol{Y}_n$.

PROOF. See Section S-3.4 of the supplement. The proof depends upon new ideas related to breaking up of relevant integrals and several lemmas, which are discussed and stated below. □

The remarks made in connection with Theorem 6.1 associated with the EW model are also applicable to Theorem 7.1 in connection with the SB model. In particular, we will choose ϵ_n, α, M_n such that the right hand side of (15) goes to zero.

The proof of Theorem 7.1 requires us to introduce some necessary concepts and technicalities. These new ideas are needed for the SB model and not for the EW model since the latter is a much less complex model than the former. In particular, note that unlike the EW case where each θ_i is represented in $L(\Theta_{M_n}, \boldsymbol{Y}_n, z)$, θ_i in the SB model may or may not be allocated to Y_i for some i, that is, there can exist z such that $z_l \neq i$, $l = 1,\ldots,n$. Suppose that $R_1^* = \{z : \text{no } \ z_l = i\}$, $R_2^* = \{z : \text{at least one } \ z_l = i\}$. Note that #$R_1^* = (M_n-1)^n$ and #$R_2^* = M_n^n - (M_n-1)^n$.

If $z \in R_1^*$, let Θ_z denote the set of θ_l's present in the likelihood and let #Θ_z=j, where $j = 1,\ldots,(M_n-1)$. By the definition of R_1^*, θ_i is not present in the likelihood. Without loss of generality let us assume that $\theta_1,\ldots,\theta_j$ are represented in the likelihood $L(\Theta_{M_n}, z, \boldsymbol{Y}_n)$. For obtaining bounds of $L(\Theta_{M_n}, z, \boldsymbol{Y}_n)$ it is enough to consider only Θ_z. For $z \in R_1^*$, we split the range of integration in the

numerator in the following way:

$$
\begin{aligned}
&\int_{\Theta_z} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}) \\
= & \int\int_{\theta_1 \in [-a-c, a+c]^c} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}) \\
& + \int\int_{\theta_1 \in [-a-c, a+c]} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}) \\
= & \int\int_{\theta_1 \in [-a-c, a+c]^c} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}) \\
& + \int\int_{\theta_1 \in [-a-c, a+c]} \int_{\theta_2 \in [-a-c, a+c]^c} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}) \\
& + \int\int_{\theta_1 \in [-a-c, a+c]} \int_{\theta_2 \in [-a-c, a+c]} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}) \\
= & \cdots \\
= & \sum_{l=1}^{j} \int_{W_l} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}) \\
& + \int_{W_{j^c}} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}),
\end{aligned}
\tag{16}
$$

where W_1=$\{\theta_1 \in [-a-c, a+c]^c$, rest θ_l's are in $(-\infty, \infty)\}$, W_l=$\{\theta_1 \in [-a-c, a+c], \ldots, \theta_{l-1} \in [-a-c, a+c], \theta_l \in [-a-c, a+c]^c$, rest θ_l's are in $(-\infty, \infty)\}$ for $l = 2, \ldots, j$, W_{j^c}=$\{\theta_1 \in [-a-c, a+c], \ldots, \theta_{j-1} \in [-a-c, a+c], \theta_j \in [-a-c, a+c]$, rest θ_l's are in $(-\infty, \infty)\}$.

Also define V_j and E as the following:

$V_j = \{z \in R_1^* : \text{ exactly } j \text{ many } \theta_l\text{'s are in } L(\Theta_{M_n}, z, y)\}$, and

E = {all θ_l's present in the likelihood are in $[-a-c, a+c]$ and the rest θ_l's are in $(-\infty, \infty)$}.

With the above developments we now state the following results which contribute to the proof of Theorem 7.1.

LEMMA 7.2. *Under the same assumptions as in Lemma 6.2 and Lemma 6.3,* $P(\sigma > \sigma_n | \boldsymbol{Y}_n) = O(\epsilon^*_{M_n})$, *where* $\epsilon^*_{M_n} = \frac{\epsilon_n}{1-\epsilon_n} \exp\left(\frac{n(a+c_1)^2}{2(b_n)^2}\right) \frac{(\alpha+M_n)^{M_n}}{\alpha^{M_n} H_0^{M_n}}$.

PROOF. See Section S-3.1 of the supplement. □

LEMMA 7.3. *Under the same assumptions of lemma 6.3,* $P(Z \in R_1^*, \Theta_{M_n} \in E^c, \sigma \leq \sigma_n | \boldsymbol{Y}_n) = O((M_n - 1)B_{M_n})$, *where* $B_{M_n} = \frac{(\alpha+M_n)}{\alpha} e^{\left(-\frac{c^2}{4\sigma_n^2}\right)}$.

PROOF. See Section S-3.2 of the supplement. □

LEMMA 7.4. *Let*

$$C_n \gtrsim \frac{n\left[\log\left(\frac{1}{\sigma_n}\right) + O\left(\frac{1}{n}\log\left(\frac{1-\epsilon_n}{\epsilon_n}\right)\right)\right]}{\left(\frac{1}{\sigma_n^2}\right)}, \tag{17}$$

where "$\gtrsim$" stands for "$\geq$" as $n \to \infty$.

Then,

$$P(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n | \mathbf{Y}_n) = O\left(\left(1 - \frac{1}{M_n}\right)^n \left(\frac{\alpha + M_n}{\alpha}\right)^{M_n}\right). \tag{18}$$

PROOF. See Section S-3.3 of the supplement. □

Note that if $M_n < \sqrt{n}$, then it is easy to verify, using L'Hospital's rule, that $\left(1 - \frac{1}{M_n}\right)^n \left(\frac{\alpha+M_n}{\alpha}\right)^{M_n} \to 0$. Lemma 7.4 thus formalizes the fact that, if the maximum number of components is small compared to the data size, then, given an appropriate estimator of σ, the probability that any mixture component will remain empty tends to zero as data size increases. On the other hand, as we show later in Lemma 13.2, if $M_n > n$, the probability that a mixture component will remain empty may converge to 1 as $n \to \infty$.

Remark: As shown in Section 3, C_n approaches the true population variance as $n \to \infty$. The condition of lemma 7.4 says that C_n is asymptotically larger than a large number. But the prior assumption on σ, the common variance of individual mixture components, tells that as $n \to \infty$, the distribution of σ becomes degenerate at 0. These two conditions on C_n and σ may appear contradictory. The answer to this apparent contradiction is that the concept of σ, the common variance of individual mixture components, is nothing to do with population variance. C_n becomes close to population variance in the long run, not to σ. Thus one should not confuse the concept of C_n with the concept of σ.

LEMMA 7.5. *$P(Z \in (R_1^*)^c, \theta_i \in [-a-c, a+c]^c, \sigma \leq \sigma_n | \boldsymbol{Y}_n) = O(B_{M_n})$, where B_{M_n} is defined in Lemma 7.3.*

PROOF. When $z \in (R_1^*)^c$, then θ_i is present in the likelihood and hence in Θ_z (defined in Lemma 7.4). Thus the same calculations associated with Lemma 7.3, now only with θ_i, guarantees the result. □

The following result asserts that the density estimators of both EW and SB converge to the same model, so that $\mu^*(y) = \theta^*(y)$ for every y.

THEOREM 7.6. *The models of EW and SB converge to the same distribution. In other words, for every y, $\mu^*(y) = \theta^*(y)$, where $\mu^*(y)$ is given in Theorem 6.1 and $\theta^*(y)$ is given in Theorem 7.1.*

PROOF. See Section S-3.5 of the supplement. □

We strengthen our convergence results given by Theorems 6.1 and 7.1 by obtaining the orders of the $MISE$ of the density estimators given by (1) and (2). In fact, Theorems 6.1 and 7.1 are directly related to the bias of the $MISE$ which can be broken up into a variance part and a bias part.

8. MISE bounds for the EW model.

8.1. *The main result for the EW model.* The main result of this section is given by the following theorem.

THEOREM 8.1. *Under the assumptions stated in Sections 2, 4, and 5,*

$$E_0^n \left[MISE(EW) \mathbb{I}_{\boldsymbol{S}_n} \right] = O\left(\left(\frac{\alpha}{\alpha+n} \right)^2 + B_n + \epsilon_n^* + \sigma_n^2 \right), \tag{19}$$

To prove Theorem 8.1 we will break up $MISE$ into variance and bias parts, following the representation (8) of $MISE$, and will obtain bounds for the variance and the bias parts separately, when $\boldsymbol{Y}_n \in \boldsymbol{S}_n$. These bounds will be independent of both y and $\boldsymbol{Y}_n$.

Note that

$$Var(\hat{f}_{EW}(y \mid \Theta_n, \sigma) | \boldsymbol{Y}_n) = \frac{1}{(\alpha+n)^2} \left[\sum_{i=1}^{n} Var\left(\frac{1}{(\sigma+k)} \phi\left(\frac{y-\theta_i}{\sigma+k} \right) \middle| \boldsymbol{Y}_n \right) \right.$$
$$\left. + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} Cov\left(\frac{1}{(\sigma+k)} \phi\left(\frac{y-\theta_i}{\sigma+k} \right), \frac{1}{(\sigma+k)} \phi\left(\frac{y-\theta_j}{\sigma+k} \right) \middle| \boldsymbol{Y}_n \right) \right]. \tag{20}$$

8.2. *Order of* $Bias(\hat{f}_{EW}(y \mid \Theta_n, \sigma))$. It follows from Theorem 6.1 that

$$\begin{aligned} & E\left(\hat{f}_{EW}(y \mid \Theta_n, \sigma) \middle| \boldsymbol{Y}_n\right) - \frac{1}{k}\phi\left(\frac{y - \mu^*(y)}{k}\right) \\ (21) \quad & = O\left(\frac{\alpha}{\alpha + n} + \frac{n}{\alpha + n}(B_n + \epsilon_n^* + \sigma_n)\right). \end{aligned}$$

For rest of this paper we denote $\frac{\alpha}{\alpha+n} + \frac{n}{\alpha+n}(B_n + \epsilon_n^* + \sigma_n)$ by S_n^*.

8.3. *Order of* $Var\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) \middle| \boldsymbol{Y}_n\right)$.

LEMMA 8.2.

$$(22) \qquad Var\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) \middle| \boldsymbol{Y}_n\right) = O\left(B_n + \epsilon_n^*\right).$$

PROOF. See Section S-4.1 of the supplement. □

8.4. *Order of the covariance term.* For $i \neq j$, let

$\xi_{in} = \left[\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) - E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) \middle| \boldsymbol{Y}_n\right)\right]$ and

$\xi_{jn} = \left[\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right) - E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right) \middle| \boldsymbol{Y}_n\right)\right]$.

$$\begin{aligned} & cov_{ij} \\ = \quad & Cov\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right), \frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right) \middle| \boldsymbol{Y}_n\right) \\ = \quad & E\left(\left[\xi_{in} - E\left(\xi_{in}\middle|\boldsymbol{Y}_n\right)\right]\left[\xi_{jn} - E\left(\xi_{jn}\middle|\boldsymbol{Y}_n\right)\right] \middle| \boldsymbol{Y}_n\right). \end{aligned}$$

LEMMA 8.3.

$$(23) \qquad cov_{ij} = O\left(B_n + \epsilon_n^*\right).$$

PROOF. See Section S-4.2 of the supplement. □

8.5. *Final calculations putting together the above results.* For $\boldsymbol{Y}_n \in \boldsymbol{S}_n$, we thus have,

$$\begin{aligned} & MISE(EW) \\ = \quad & O\left(\frac{1}{(\alpha+n)^2}\left[n(B_n + \epsilon_n^*) + n(n-1)(B_n + \epsilon_n^*)\right] + (S_n^*)^2\right). \end{aligned} \qquad (24)$$

Assuming n to be large enough such that $\frac{n(n-1)}{(\alpha+n)^2} \approx 1$, the actual form of $MISE$ given in equation (24) can be simplified further for comparison purpose. Note that if $\frac{\alpha}{\alpha+n} \to 1$, then the conditional distribution based on the Polya urn scheme implies that θ_l's arise from G_0 only, which seems to be too restrictive an assumption. Thus assuming $\frac{\alpha}{\alpha+n} \to 0$ seems more plausible as it entails a nonparametric set up. We assume $\alpha = n^{\omega}$, $\omega < 1$ and assume that, $\frac{n}{\alpha+n} \approx 1$ for large n. So, (24) boils down to

$$\begin{aligned} & MISE(EW) \\ = \; & O\left(\frac{n}{(\alpha+n)^2}(B_n+\epsilon_n^*) + (B_n+\epsilon_n^*) + \left(\frac{\alpha}{\alpha+n} + B_n + \epsilon_n^* + \sigma_n\right)^2\right), \end{aligned} \tag{25}$$

for $\boldsymbol{Y}_n \in \boldsymbol{S}_n$. We can further simplify this form by retaining only the higher order terms. Note that we have assumed that under certain conditions B_n, ϵ_n^* and $\frac{\alpha}{\alpha+n}$ converge to 0, and hence $\frac{n}{(\alpha+n)^2}(B_n+\epsilon_n^*) \prec O\left(B_n+\epsilon_n^*\right)$. In the third term of equation (25) there are two extra terms, σ_n and $\frac{\alpha}{\alpha+n}$ under the squared term. Adjusting for that term we write the simplified form of of $MISE$, for $\boldsymbol{Y}_n \in \boldsymbol{S}_n$, as

$$MISE(EW) = O\left(\left(\frac{\alpha}{\alpha+n}\right)^2 + B_n + \epsilon_n^* + \sigma_n^2\right). \tag{26}$$

Note that the order remains unchanged after multiplication with $\mathbb{I}_{\boldsymbol{S}_n}$ and taking expectation with respect to E_0^n. In other words, Theorem 8.1 follows.

9. $MISE$ bounds for the SB model.

9.1. *The main result for the SB model.*

THEOREM 9.1. *Under the above assumptions stated in Sections 3, 4, and 5,*

$$E_0^n\left[MISE(SB)\mathbb{I}_{\boldsymbol{S}_n}\right] = O\left(\left(1-\frac{1}{M}\right)^n\left(\frac{\alpha+M}{\alpha}\right)^M + MB_M + \epsilon_M^* + \sigma_n^2\right).$$

9.2. *Bounds of* $Var(\hat{f}_{SB}(y \mid \Theta_M, \sigma))$.

LEMMA 9.2.

$$\begin{aligned} & \frac{1}{M^2}\sum_{i=1}^{M} Var\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right) \\ = \; & O\left(\frac{1}{M}\left(MB_M + \left(1-\frac{1}{M}\right)^n\left(\frac{\alpha+M}{\alpha}\right)^M + \epsilon_M^*\right)\right) \end{aligned} \tag{27}$$

PROOF. See Section S-5.1 of the supplement. □

9.3. *Order of the covariance term.*

LEMMA 9.3.

$$\sum_{i=1}^{M}\sum_{j=1,j\neq i} Cov\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right),\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right)$$

$$= O\left(MB_M+\left(1-\frac{1}{M}\right)^n\left(\frac{\alpha+M}{\alpha}\right)^M+\epsilon_M^*\right) \tag{28}$$

PROOF. See Section 5.2 of the supplement. □

9.4. *Bound for the bias term.* The bias of the $MISE$, which we denote by $Bias(\hat{f}_{SB}(y \mid \Theta_M,\sigma))$, is given by

$$\frac{1}{M}\sum_{j=1}^{M}E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right)-\frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right).$$

From Theorem 7.1 we have,

$$Bias(\hat{f}_{SB}(y \mid \Theta_M,\sigma))^2 = O\left(\left[MB_M+\left(1-\frac{1}{M}\right)^n\left(\frac{\alpha+M}{\alpha}\right)^M+\epsilon_M^*+\sigma_n\right]^2\right). \tag{29}$$

Thus, the complete order of $MISE$ can be obtained by adding up these individual orders of (27), (28) and (29), yielding

$$\begin{aligned} MISE(SB) = {} & O\left(\frac{1}{M}\left(MB_M+\left(1-\frac{1}{M}\right)^n\left(\frac{\alpha+M}{\alpha}\right)^M+\epsilon_M^*\right)\right)\\ & +O\left(MB_M+\left(1-\frac{1}{M}\right)^n\left(\frac{\alpha+M}{\alpha}\right)^M+\epsilon_M^*\right)\\ & +O\left(\left[MB_M+\left(1-\frac{1}{M}\right)^n\left(\frac{\alpha+M}{\alpha}\right)^M+\epsilon_M^*+\sigma_n\right]^2\right) \end{aligned} \tag{30}$$

Appropriate choices of the sequences involved in $MISE(SB)$ guarantee that $MB_M \to 0$, $\left(1-\frac{1}{M}\right)^n \left(\frac{\alpha+M}{\alpha}\right)^M \to 0$ and $\epsilon^*_M \to 0$. With these we have

$$MISE(SB) \quad = \quad O\left(\left(1-\frac{1}{M}\right)^n \left(\frac{\alpha+M}{\alpha}\right)^M + MB_M + \epsilon^*_M + \sigma_n^2\right). \tag{31}$$

Note that the order remains unchanged after multiplication with $\mathbb{I}_{\boldsymbol{S}_n}$ and taking expectation with respect to E_0^n, thus proving Theorem 9.1.

10. Comparison between MISE's of EW and SB. As claimed in Mukhopadhyay, Bhattacharya and Dihidar (2011) and Mukhopadhyay, Roy and Bhattacharya (2012), the SB model is much more efficient than the EW model in terms of computational complexity and ability to approximate the true underlying clustering or regression. Here we investigate the conditions under which the model of SB beats that of EW in terms of $MISE$. In particular, we provide conditions which guarantee that each term of the order of $MISE(EW)$ dominates the corresponding term of the order of $MISE(SB)$ (for any two sequences $\{a_n^{(1)}\}$ and $\{a_n^{(2)}\}$ we say that $a_n^{(1)}$ dominates $a_n^{(2)}$ if $a_n^{(2)}/a_n^{(1)} \to 0$ as $n \to \infty$).

For the purpose of comparison we will use the same values of b_n, σ_n, ϵ_n, for all n, for both SB and EW model, in a way such that both the $MISE$'s converge to 0.

LEMMA 10.1. *Let $\alpha = n^\omega$, $M = n^b$, where $\omega < 1$, $b < 1$, and $\omega < b$. Then,* $\frac{\epsilon^*_M}{\epsilon^*_n} \to 0$.

PROOF. The proof follows from simple applications of L'Hospital's rule. □

LEMMA 10.2. *Let $M = n^b$ and $\alpha = n^\omega$, where $b > \omega$. Then $B_n > MB_M$ if $M < \sqrt{n}$.*

PROOF. The proof follows from simple applications of L'Hospital's rule. □

LEMMA 10.3. $r_1(n) = \dfrac{\left(1-\frac{1}{M}\right)^n \left(\frac{\alpha+M}{\alpha}\right)^M}{\left(\frac{\alpha}{\alpha+n}\right)^2} \to 0$, *if* $\frac{1}{2} > b > \omega$.

PROOF. The proof follows from simple applications of L'Hospital's rule. □

Hence, combining the results of Lemma 10.1 to Lemma 10.3 we conclude that $MISE(SB)$ converges to 0 at a faster rate than $MISE(EW)$, provided that we choose M and α as required by Lemmas 10.1–10.3.

11. Comparison with the optimal frequentist $MISE$ rate. Under suitable regularity conditions, the optimal $MISE$ rate of convergence of a kernel density estimator is $n^{-2/5}$. Ghosal and van der Vaart (2007) consider densities of the form

$$f_F(x) = \int \phi\left(\frac{x-\theta}{\sigma}\right) dF(\theta), \tag{32}$$

where $F \sim D(\alpha G_0)$. Thus, in contrast to our approach where we considered predictive densities integrating out the random measure F, Ghosal and van der Vaart (2007) deal with densities of the form (32) whose randomness is induced by F. The remaining prior structure is very similar to ours. Assuming that f_0 is the true density generating the data, Ghosal and van der Vaart (2007) consider convergence (to zero) under the true distribution of the data, of posterior probabilities of the form $P\left(F : d_h(f_F, f_0) > C\epsilon_{0,n} \mid \boldsymbol{Y}_n\right)$, where $\epsilon_{0,n}$ is the rate of convergence that Ghosal and van der Vaart (2007) are interested in, C is some constant and d_h is some pseudo-metric; in particular, Ghosal and van der Vaart (2007) consider the Hellinger distance between two densities. Under certain assumptions the best rate of convergence turns to be $n^{-2/5}(\log n)^{4/5}$, which is slower than the frequentist $MISE$ rate.

Here we show that in our approach, by choosing α, ϵ_n and σ_n appropriately, $MISE(SB)$ can achieve a smaller rate than $n^{-2/5}$. Moreover, with our approach, $MISE(EW)$ can achieve the optimal frequentist rate $n^{-2/5}$.

Let $\alpha = n^{\omega}$, $0 < \omega < 1$. We set $\sigma_n^2 = \frac{1}{n^t}$, $t > 0$ and ϵ_n is chosen so that $\epsilon_n^* = \frac{1}{n^r}$, $r > 0$. Then the order of $MISE(EW)$ becomes

$$\frac{1}{(1+n^{1-\omega})^2} + \left(\frac{\alpha+n}{\alpha} e^{-\frac{c^2 n^t}{4}}\right) + \frac{1}{n^r} + \frac{1}{n^t}. \tag{33}$$

If we choose $r < \frac{2}{5}$, $t < \frac{2}{5}$ and $\omega < \frac{4}{5}$, then (33) will be less than $n^{-2/5}$. Moreover, under the same conditions $MISE(SB)$ has a smaller rate than $MISE(EW)$. The form of our true distribution satisfies the assumption of Ghosal and van der Vaart (2007) which requires it to tend to zero smoothly at the boundary points of its support; it also satisfies those required for frequentist density estimation. Hence, for our form of the true distribution, the convergence rates of the different methods are comparable–we obtain the following ordering: $MISE(SB) \prec MISE(EW) \prec FMISE \prec BR_{GVV}$, where $FMISE$ denotes the optimal frequentist $MISE$ rate and BR_{GVV} denotes the best rate of convergence of Ghosal and van der Vaart (2007).

We now show that for $\sigma_n = n^{-1/5}$ (which can be thought of as analogous to the optimum bandwidth in the frequentist density estimation paradigm) $MISE(EW)$ can achieve the rate $n^{-2/5}$ when minimized with respect to α over $0 \leq \alpha <$

∞, that is, without assuming any particular form of α. Taking $\sigma_n^2 = n^{-t}$, the order of $MISE(EW)$ is minimized for $\alpha = n\left(\frac{1}{1-\frac{e^{-\frac{c^2 n^t}{12}}}{2^{1/3}}} - 1\right)$. The order of $MISE(EW)$ in that case becomes

$$\left(\frac{1}{2^{1/3}} e^{-\frac{c^2 n^t}{12}}\right)^2 + 2^{1/3} e^{c^2} + \frac{1}{n^r} + \frac{1}{n^t}. \tag{34}$$

Hence, if $\sigma_n^2 = \frac{1}{n^{2/5}}$, then the order of $MISE(EW)$ is also $\frac{1}{n^{2/5}}$, if $r > t$.

However, note that the optimum order of $MISE(EW)$ given by (34) is attained when α is a decreasing function of n. This is not the same condition under which we have proved better performance of the SB model over the EW model. So it is of interest to study whether under this new condition also the SB model outperforms the EW model. Using L'Hospital's rule we can show, putting $\alpha = n\left(\frac{1}{1-\frac{e^{-\frac{c^2 n^t}{12}}}{2^{1/3}}} - 1\right)$ in Lemmas 10.1–10.3, that the corresponding ratios still converge to 0 as $n \to \infty$. Thus the SB model can achieve even smaller optimal posterior $MISE$ rate compared to the other methods.

12. The "large p small n" problem.

So far we have discussed the asymptotic performance of SB and EW models in terms of $MISE$ for univariate data. We now investigate the $MISE$-based convergence properties when the sample size n is much smaller than p, the dimension of each data point. In rough terms, the information contained in the data is much less compared to the number of parameters to be estimated, which makes inference extremely challenging. This problem is the well-known "large p small n" problem. We assume that as $n \to \infty$, $p \to \infty$ such that $p/n \to \infty$. In fact, the data dimension should more appropriately be denoted by p_n, but for the sake of convenience we suppress the suffix.

12.1. *EW model.*

The EW model for p variate data is defined as follows.

$$Y_i \sim N_p(\theta_i, \Sigma) \quad \text{independently}, \tag{35}$$

where $Y_i = (Y_{i1}, \ldots, Y_{ip})'$, $\theta_i = (\theta_{i1}, \ldots, \theta_{ip})'$, $i = 1, \ldots, n$ and $\Sigma = \sigma^2 I_{p\times p}$, where $I_{p\times p}$ is the $p \times p$ identity matrix.

$$\theta_i \overset{ind}{\sim} F, i = 1, \ldots, n; F \sim D(\alpha G_0), \tag{36}$$

where under G_0, $\theta_i \overset{ind}{\sim} N_p(\mu_0, \Sigma_0)$, where $\Sigma_0 = \sigma_0^2 I_{p\times p}$; μ_0 and σ_0 are assumed to be known. The prior on σ remains the same as before. The predictive density at the point $y = \{y_1, \ldots, y_p\}$ is given by:

$$\hat{f}_{EW}(y \mid \Theta_n, \sigma) = \frac{\alpha}{\alpha+n} A_n + \frac{1}{\alpha+n} \sum_{i=1}^{n} \prod_{l=1}^{p} \frac{1}{(\sigma+k)} \phi\left(\frac{y_l - \theta_{il}}{\sigma+k}\right). \tag{37}$$

Let $\boldsymbol{S}_n = \{\boldsymbol{Y}_n : |Y_{ij}| < a; i = 1, \ldots, n, j = 1, \ldots, p\}$. The form of the $MISE$ and the approach to bound the $MISE$ will be same as in Section 2. So, we proceed directly to obtain the bounds of the different parts of the $MISE$.

LEMMA 12.1. *Under the same assumptions as in Lemma 6.2,* $P(\sigma > \sigma_n \mid \boldsymbol{Y}_n) = O(\epsilon_n^L)$, *where* $\epsilon_n^L = \frac{\epsilon_n}{1-\epsilon_n} \exp\left(\frac{np(a+c_1)^2}{2b_n^2}\right) \frac{(\alpha+n)^n}{(\alpha)^n H_0^{np}}$.

PROOF. See Section S-6.1 of the supplement. □

Let E=$\{\theta_{il} \in [-a-c, a+c], \forall l$, rest θ_l's are in $\Re^p\}$ and E^c=$\{$at least one $\theta_{il} \in [-a-c, a+c]^c, l = 1(1)p$, rest θ_l's are in $\Re^p\}$; $\Re$ representing the real line. We then have the following result.

LEMMA 12.2. *Under the conditions of Lemma 6.3 it can be shown that* $P(\Theta_n \in E^c, \sigma \leq \sigma_n \mid \boldsymbol{Y}_n) = O(pB_n)$ *where* $B_n = \frac{\alpha+n}{\alpha} e^{-\frac{c^2}{4\sigma_n^2}}$.

PROOF. See Section S-6.3 of the supplement. The proof proceeds by splitting the integral $\int_{\Theta_n} \int_0^{\sigma_n} L(\Theta_n, \boldsymbol{Y}_n) dH(\Theta_n) dG_n(\sigma)$ to sum of integrals over particular regions provided in Section S-6.2 of the supplement. □

Hence, from Lemmas 12.1 and 12.2, and from the same argument used for obtaining the order of $MISE(EW)$ in Section 2, here we obtain the order of $MISE(EW)$ in the large p small n case as

$$E_0^n \left[MISE(EW)\mathbb{I}_{\boldsymbol{S}_n}\right] = O\left(\left(\frac{\alpha}{\alpha+n}\right)^2 + pB_n + \epsilon_n^L + \sigma_n^2\right). \tag{38}$$

12.2. *SB model.* Here the model is

$$\hat{f}_{SB}(y \mid \Theta_M, \sigma) = \frac{1}{M} \sum_{i=1}^{M} \prod_{l=1}^{p} \frac{1}{(\sigma+k)} \phi\left(\frac{y_l - \theta_{il}}{\sigma+k}\right), \tag{39}$$

where $\Theta_M = (\theta_1, \ldots, \theta_M)'$, $\theta_i = (\theta_{i1}, \ldots, \theta_{ip})'$, $\Sigma = \sigma^2 I_{p\times p}$ and the prior assumptions are the same as in equations (35) and (36).

We bound the individual probabilities in the same way as in Section 9.

LEMMA 12.3. *$P\left(\sigma \geq \sigma_n \mid \boldsymbol{Y}_n\right) = O\left(\epsilon_M^L\right)$, where* $\epsilon_M^L = \frac{\epsilon_n}{1-\epsilon_n}\exp\left(\frac{np(a+c_1)^2}{2(b_n)^2}\right)\frac{(\alpha+M)^M}{(\alpha)^M H_0^{Mp}}$.

PROOF. Here the form of the likelihood is

$$L\left(\Theta_M, z, \boldsymbol{Y}_n\right) = \frac{1}{\sigma^{np}} e^{-\sum_{j=1}^{M}\sum_{t:z_t=j}\sum_{l=1}^{p}\frac{(y_{tl}-\theta_{tl})^2}{2\sigma^2}} \tag{40}$$

Application of same technique as in Lemma 7.2 leads to the required result. □

Let E ={all θ_l's in likelihood are in $[-a-c, a+c]$, rest are in $\Re^p$}.

LEMMA 12.4. *$P(Z \in R_1^*, \Theta_M \in E^c, \sigma \leq \sigma_n) = O((M-1)pB_M)$, B_M same as in Lemma 7.3.*

PROOF. In the same way as in Lemma 7.3 it follows that

$$P(Z \in R_1^*, \Theta_M \in E^c, \sigma \leq \sigma_n) \leq (M-1)pB_M.$$

Splitting the integral $\int_0^{\sigma_n}\int_{\Theta_M} L\left(\Theta_M, z, \boldsymbol{Y}_n\right) dH(\Theta_n) dG_n(\sigma)$ required for this proof is provided in Section S-7. □

LEMMA 12.5. *$P(Z \in R_1^*, \Theta_M \in E, \sigma \leq \sigma_n) = O\left(\left(1-\frac{1}{M}\right)^n\left(\frac{\alpha+M}{\alpha}\right)^M\right)$.*

PROOF. Exactly same as the proof of Lemma 7.4. □

LEMMA 12.6. *$P\left(Z \in (R_1^*)^c, \Theta_M \in E^c, \sigma \leq \sigma_n\right) = O\left(pB_M\right)$.*

PROOF. The proof follows in the same way as that of Lemma 7.3. □

Proceeding in the same way as in Section 9 we can show that the order of $MISE$ of SB model for the "large p small n" case is

$$E_0^n\left[MISE(SB)\mathbb{I}_{\boldsymbol{S}_n}\right] = O\left(\left(1-\frac{1}{M}\right)^n\left(\frac{\alpha+M}{\alpha}\right)^M + MpB_M + \epsilon_M^L + \sigma_n^2\right). \tag{41}$$

Note that the two "large p, small n" $MISE$'s given by (38) and (41) converge to zero if σ_n and ϵ_n converge to zero at a much faster rate than the rate needed for $MISE$'s given by (26) and (31). This is clear because p grows much faster than n and so, the terms pB_n and MpB_M involved in the $MISE$'s do not go to zero unless σ_n is made to converge to zero much faster than in the usual cases. Also, ϵ_n must go to zero much faster than in the usual situations to ensure that ϵ_n^L and ϵ_M^L converge to zero. In other words, rather strong prior information is necessary to compensate for the deficiency of information in the data.

The above arguments show that even in the "large p small n" problem both EW and SB model can be consistent in terms of $MISE$ by choosing the prior on σ appropriately. However, it is easy to see that even in this set-up, $MISE(SB) \prec MISE(EW)$.

13. Convergence to wrong model.

We now investigate conditions under which the models of EW and SB converge to wrong models, that is, to models which did not generate the data. It is perhaps easy to anticipate that if α is made to grow at a rate faster than n, then the density estimator of EW would converge to the convolution of the kernel and G_0, irrespective of the true, data-generating model. If G_0 has non-compact support then the convolution can not be represented in the form $\frac{1}{k}e^{-\frac{(y-\theta^*)^2}{2k^2}}$, which we have assumed to be the form of the true distribution from which the data are obtained. We show in this section that indeed the simple condition of letting α grow faster than n is enough to derail the EW model. On the other hand, much stronger conditions are necessary to get the SB model to converge to the wrong model.

13.1. *EW model.*

THEOREM 13.1. *Suppose that $\alpha \succ O(n)$. Then*

$$E_0^n\left[E\left(\hat{f}_{EW}(y \mid \Theta_n, \sigma)\right)\mathbb{I}_{\boldsymbol{S}_n}\right] \rightarrow \int_{\theta_{n+1}} \frac{1}{k}e^{-\frac{(y-\theta_{n+1})^2}{2k^2}}dG_0(\theta_{n+1}). \tag{42}$$

PROOF. See Section S-8.1 of the supplement. □

Thus, the condition $\alpha \succ O(n)$ is enough to mislead the EW model, taking it to a wrong model.

13.2. *SB model.*

Recall that no condition on α is necessary to ensure asymptotic convergence of $MISE(SB)$ given by (31); only σ_n needs to be chosen appropriately and that $M < \sqrt{n}$. So, unlike in the case of EW, even if $\alpha \succ O(M)$

or $\alpha \succ O(n)$, the SB model can still converge to the true distribution by setting $M \prec O(\sqrt{n})$.

However, for $M \succ O(n)$, the term $\left(1-\frac{1}{M}\right)^n \left(\frac{\alpha+M}{\alpha}\right)^M$ does not converge to 0. This term is associated with the upper bound of the posterior probability $P(Z \in R_1^*, \Theta_M \in E, \sigma \leq \sigma_n|\boldsymbol{Y}_n)$. The next question arises whether there are conditions under which this probabilty can converge to 1 as $n \to \infty$. Towards answering this question, let us consider the following result.

LEMMA 13.2. *Let $\{r_n\}$ be a sequence tending to zero such that $O\left(\log\left(\frac{1}{\epsilon_n}\right)\right) \prec -n\log(r_n) - n\log(n)$, and let $C_n = O\left(\frac{1}{r_n^s n^2}\right)$; $s > 2$. Then*

$$P(Z \in R_1^*, \Theta_M \in E, \sigma \leq \sigma_n|\boldsymbol{Y}_n) \gtrsim \left(1-\frac{1}{M}\right)^n \left(\frac{\alpha+M}{\alpha}\right)^M. \tag{43}$$

PROOF. See Section S-8.2 of the supplement. □

Remark 1: Note that $C_n = O\left(\frac{1}{r_n^s n^2}\right)$; $s > 2$, as required by Lemma 13.2 need not necessarily ensure that $\frac{C_n}{n}$ remains bounded as $n \to \infty$. In such cases we can not assume that $|Y_i| < a; i = 1,\ldots,n$, since this forces $\frac{C_n}{n}$ to be bounded. Thus, we must assume in such cases that $-\infty < Y_i < \infty; i = 1,\ldots,n$. This has no effect on the proofs of our results in this section since none of the proofs of these results depends upon the range of the Y_i's.

Remark 2: If Y_i are not bounded, for $M \succ O(n)$, it is expected that $\frac{C_n}{n}$ will grow with n. This is because the mixture components from which the data arise may be very widely separated when the number of mixture components far exceeds the number of data points.

Remark 3 : Remarks 1 and 2 above shows that for the probability (43) to be large, it may require the data to be come from unbounded regions, and to have very high variability. Since Lemma 13.2 and the probability given by (43) are instrumental for the convergence of the SB model to the wrong model (to be seen subsequently), it follows that in this respect as well SB model seems to be superior to the EW model, since the latter does not require similar assumptions on unboundedness.

Using L' Hospital's Rule it is easy to check that for $M = n^b$, $\alpha = n^\omega$, $\omega > 0$, $b > 0$, $\omega, b > 1$, and $\omega - b > b$, $\left(\frac{\alpha}{\alpha+M}\right)^M \left(1-\frac{1}{M}\right)^n \to 1$. That is, if $M > n$, the probability that a mixture component will remain empty may converge to 1 as

$n \to \infty$.

Thus for $M > n$, $\alpha > n$, the posterior probability of $\{Z \in R_1^*, \Theta_M \in E, \sigma \leq \sigma_n\}$ converges to 1, as $n \to \infty$. Hence, in $E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right)$, the factor associated with the above posterior probability is the only contributing term for large n. We now study convergence of this term.

THEOREM 13.3. *Assume the conditions of Lemma 13.2. Further assume that* $M = n^b$, $\alpha = n^\omega$, $\omega > 1$, $b > 1$ *and* $\omega - b > b$. *Then, for the SB model it holds that*

$$E_0^n\left[E\left(\hat{f}_{SB}(y \mid \Theta_M, \sigma)\middle|\boldsymbol{Y}_n\right)\mathbb{I}_{\boldsymbol{S}_n}\right] \to \int_{\theta_i}\frac{1}{k}e^{-\frac{(y-\theta_i)^2}{2k^2}}dG_0(\theta_i). \tag{44}$$

PROOF. See Section S-8.3 of the supplement. □

Thus, while the EW model can converge to the wrong model if only $\alpha \succ O(n)$ is assumed, much stronger restrictions are necessary to get the SB model to deviate from the true model.

14. Modified SB Model. A slightly modified version of SB model is as follows:

$$\hat{f}_{SB}^*(y \mid \Theta_M, \Pi, \sigma) = \sum_{i=1}^{M}\pi_i\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right), \tag{45}$$

where $\sum_{i=1}^M \pi_i = 1$. We assume that $\Pi = (\pi_1, \ldots, \pi_M) \sim Dirichlet(\beta_1, \ldots, \beta_M)$, $\beta_i > 0$ and is independent of Θ_M and σ. The assumptions of Dirichlet process prior on Θ_M and the prior structure of σ remain same as before. The previous form of the SB model (2) is a special case (discrete version) of this model with $\pi_i = \frac{1}{M}$ for each i.

Due to discreteness of the Dirichlet process prior, the parameters θ_i are coincident with positive probability. As a result, (45) reduces to the form

$$\hat{f}_{SB}^*(y \mid \Theta_M, \Pi, \sigma) = \sum_{i=1}^{M^*}p_i\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i^*}{\sigma+k}\right), \tag{46}$$

where $\{\theta_1^*, \ldots, \theta_{M^*}^*\}$ are M^* distinct components in $\boldsymbol{\Theta}_M$ with θ_i^* occuring M_i times, and $p_i = \sum_{j=1}^{M_i}\pi_j$. In contrast to the previous form of the SB model (2)

where the mixing probabilities are of the form M_i/M, here the mixing probabilities p_i are continuous.

The asymptotic calculations associated with the modified SB model are almost the same as in the case of the SB model in Section 9. Indeed, this modified version of SB's model converges to the same distribution where the EW model and the previous version of the SB model also converge. Moreover, the order of $MISE$ for this model remains exactly the same as that of the previous version of the SB model. In Section S-9 of the supplement we provide a brief overview of the steps involved in the asymptotic calculations.

SUPPLEMENTARY MATERIAL

Throughout, we refer to our main paper Mukhopadhyay and Bhattacharya (2012a) as MB.

S-1. Proofs of results associated with Section 3 of MB.

S-1.1. *Proof of Lemma 3.1.*

PROOF. Let $C_n^{'} = \frac{\sum_{j=1}^{M_n} n_j \left(\bar{Y}_j - \bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}$. We recall that $\boldsymbol{Y}_n$ form a triangular array, as argued in Section 2 of MB. The n-th row of that array is summarized by the statistic $C_n^{'}$. Since the random variables of a particular row of that array are independent of the random variables of the other rows, C_n' are independent among themselves.

Suppose μ_T is the true population mean and σ_T^2 is the true population variance (both of which are assumed to be finite). Since all Y_i's are from same true density f_0, under f_0, $E^{f_0}(Y_i) = \mu_T$ and $V^{f_0}(Y_i) = \sigma_T^2$.

$$
\begin{aligned}
& E^{f_0}\left(\frac{\sum_{j=1}^{M_n} n_j \left(\bar{Y}_j - \bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}\right) \\
= \; & E_z\left[E^{f_0}_{Y|z}\left(\frac{\sum_{j=1}^{M_n} n_j \left(\bar{Y}_j - \bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}\right)\right]. \qquad (1)
\end{aligned}
$$

Note that

$$
\begin{aligned}
& E^{f_0}_{Y|z}\left(\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}\right) \\
= \; & \frac{1}{n}E^{f_0}_{Y|z}\left(\sum_{j=1}^{M} n_j(\bar{Y}_j-\mu_T)^2-n(\bar{Y}-\mu_T)^2\right) \\
= \; & \frac{M_n-1}{n}\sigma_T^2=\mu_n^*, \qquad (2)
\end{aligned}
$$

noting the fact that since, given that $Y\sim f_0$, Y and Z are independent, $V^{f_0}_{Y|z}(Y_i)=\sigma_T^2$. Hence,

$$
E^{f_0}\left(\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}\right)=\frac{M_n-1}{n}\sigma_T^2=\mu_n^*. \qquad (3)
$$

Note that if $M_n>1$ for all n, then $\mu_n^*>0$ and if $M_n\prec O(n)$, then $\mu_n^*\to 0$.

Similarly we can split the variance term as

$$
\begin{aligned}
& V^{f_0}\left(\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}\right) \\
= \; & V_z\left[\frac{1}{n}E^{f_0}_{Y|z}\left(\sum_{j=1}^{M} n_j(\bar{Y}_j-\mu_T)^2-n(\bar{Y}-\mu_T)^2\right)\right] \\
+ \; & E_z\left[V^{f_0}_{Y|z}\left(\frac{1}{n}\sum_{j=1}^{M} n_j(\bar{Y}_j-\mu_T)^2-n(\bar{Y}-\mu_T)^2\right)\right]. \qquad (4)
\end{aligned}
$$

From (2) we have $E^{f_0}_{Y|z}\left(\sum_{j=1}^{M} n_j(\bar{Y}_j-\mu_T)^2-n(\bar{Y}-\mu_T)^2\right)$ is free of z. So the first term in the summation of (4) is 0. Easy calculations shows that the order the second term in that summation is $\frac{2^n}{n\times(M_n)^n}$.
Thus

$$
V^{f_0}\left(\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}\right)=O\left(\frac{2^n}{n\times(M_n)^n}\right). \qquad (5)
$$

Note that for $M_n>2$, $V^{f_0}\left(\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}\right)$ converges to 0. Hence, under f_0, $\left|\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}-\mu_n^*\right|\to 0$, in probability, for $M_n>1$ and $M_n\prec O(n)$.

Also we note that, under f_0,

$$P\left(\left[\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}-\mu_n^*\right]^2>\epsilon\right) \leq \frac{1}{\epsilon}E^{f_0}\left(\left[\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}-\mu_n^*\right]^2\right)$$
$$= O\left(\frac{2^n}{n\times(M_n)^n}\right).$$

Thus, for $M_n>1$,

$$\sum_{n=1}^{\infty}P\left(\left[\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}-\mu_n^*\right]^2>\epsilon\right)<\infty.$$

Hence we conclude that, under f_0, $\left|\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}-\mu_n^*\right|\to 0$, a.s., for $M_n>1$ and $M_n\prec O(n)$. Also we have $\mu_n^*\to 0$, a.s. under the same set of conditions. Combining these two results we have that, under f_0, $\frac{\sum_{j=1}^{M_n} n_j\left(\bar{Y}_j-\bar{Y}\right)^2}{\sum_{j=1}^{M_n} n_j}\to 0$, a.s., for $M_n>1$ and $M_n\prec O(n)$.

□

S-2. Proofs of results associated with Section 6 of MB.

S-2.1. *Proof of Lemma 6.2.*

PROOF. $P(\sigma>\sigma_n|\boldsymbol{Y}_n)=\frac{\int_{\sigma_n}^{\infty}\int_{\Theta_n}\prod_{j=1}^{n}\frac{1}{\sigma}\phi\left(\frac{Y_j-\theta_j}{\sigma}\right)dG_n(\sigma)dH_n\left(\Theta_n\right)}{\int_{\sigma}\int_{\Theta_n}\prod_{j=1}^{n}\frac{1}{\sigma}\phi\left(\frac{Y_j-\theta_j}{\sigma}\right)dG_n(\sigma)dH_n\left(\Theta_n\right)}=\frac{N}{D}$, where $H_n\left(\Theta_n\right)$ is the joint distribution of Θ_n.

Denote $L(\Theta_n,Y)=\prod_{j=1}^{n}\frac{1}{\sigma}\phi\left(\frac{Y_j-\theta_j}{\sigma}\right)$.

$$\begin{aligned}
D &\geq \int_{\sigma\in(b_n,\sigma_n)}\int_{\Theta_n\in E_n}L(\Theta_n,Y)dG_n(\sigma)dH_n\left(\Theta_n\right)\\
&= \frac{e^{-\left\{\frac{\sum_{i=1}^{n}(Y_i-\theta_i^*)^2}{2{b_n}^2}\right\}}}{(b_n)^n}P(b_n<\sigma\leq\sigma_n)P\left(\Theta_n\in E_n\right),
\end{aligned}$$

where $E_n=\{\theta_j\in[-c_1,c_1],i=1,\ldots,n\}$, and $\theta_i^*\in(-c_1,c_1)$, $c_1>0$ is a very small constant.

Now $|Y_i| < a, |\theta_i^*| < c_1 \Rightarrow (Y_i - \theta_i^*)^2 < (a + c_1)^2, \forall i \Rightarrow \sum_{i=1}^n (Y_i - \theta_i^*)^2 < n(a+c_1)^2$. Again, from properties of the Polya urn, $H_n(\Theta_n) \geq \prod_{i=1}^n \frac{\alpha}{\alpha+n} G_0(\theta_i) \Rightarrow P(\Theta_n \in E_n) \geq (\frac{\alpha}{\alpha+n})^n H_0^n$.

Now, the function $\frac{1}{\sigma}\exp\{-\frac{n(a+c_1)^2}{2\sigma^2}\}$ is increasing on $b_n < \sigma < \sigma_n$ for $\sigma_n < \sqrt{n}(2a+c)$. Hence,

$$D \geq \frac{e^{\left(\frac{-n(a+c_1)^2}{2{b_n}^2}\right)}}{(b_n)^n} P(b_n < \sigma \leq \sigma_n)\left(\frac{\alpha}{\alpha+n}\right)^n H_0^n. \tag{6}$$

For the numerator observe that for $\sigma > \sigma_n$, $\dfrac{e^{-\frac{\sum_{i=1}^n (Y_i-\theta_i)^2}{2\sigma^2}}}{\sigma^n} \leq \frac{1}{(\sigma_n)^n}$. This implies

$$\begin{aligned} N = \int_{\Theta_n}\int_{\sigma>\sigma_n} L(\Theta_n, Y) dG_n(\sigma) dH_n(\Theta_n) &\leq \frac{1}{(\sigma_n)^n} P(\sigma > \sigma_n) \\ &= \frac{O(\epsilon_n)}{(\sigma_n)^n}. \end{aligned} \tag{7}$$

(6) and (7) together implies that,

$$P(\sigma > \sigma_n | \boldsymbol{Y}_n) \leq \frac{P(\sigma > \sigma_n)}{1 - P(b_n < \sigma \leq \sigma_n)} \frac{{b_n}^n}{(\sigma_n)^n} e^{\frac{n(a+c_1)^2}{2{b_n}^2}} \frac{(\alpha+n)^n}{(\alpha)^n H_0^n} = A_n^*, \text{say}.$$

Since $b_n \leq \sigma_n$, we have ${b_n}^n \leq (\sigma_n)^n$; also by the assumption of the lemma $P(\sigma > \sigma_n) = O(\epsilon_n)$, $P(\sigma < \sigma_n) = O(1-\epsilon_n)$, so that $P(b_n < \sigma < \sigma_n) \leq P(\sigma < \sigma_n)$ implies $P(b_n < \sigma < \sigma_n) = O(1-\epsilon_n)$. Thus $A_n^* = O(\epsilon_n^*)$, where $\epsilon_n^* = \frac{\epsilon_n}{1-\epsilon_n} e^{\frac{n(a+c_1)^2}{2{b_n}^2}} \frac{(\alpha+n)^n}{(\alpha)^n H_0^n}$. This completes the proof. □

S-2.2. *Proof of Lemma 6.4.*

PROOF.

$$\begin{aligned} A_n &= \int_\theta \frac{1}{(\sigma+k)} e^{-\frac{(y-\theta)^2}{2(\sigma+k)^2}} dG_0(\theta) \\ &\leq H_1 \int_\theta dG_0(\theta) \\ &= H_1. \end{aligned} \tag{8}$$

where $H_1 = \sup_{\{y,\theta,\sigma\}} \left\{ \frac{1}{(\sigma+k)} e^{-\frac{(y-\theta)^2}{2(\sigma+k)^2}} \right\} = \frac{1}{k}$. Thus $A_n = O(1)$, and

$$\frac{\alpha}{\alpha+n} A_n = O\left(\frac{\alpha}{\alpha+n}\right). \tag{9}$$

Since $\alpha = O(n^\omega); 0 < \omega < 1$, $\frac{\alpha}{\alpha+n} \to 0$, and hence the proof follows. □

S-2.3. *Proof of Theorem 6.1.*

PROOF. For $\boldsymbol{Y}_n \in \boldsymbol{S}_n$,

$$
\begin{aligned}
& E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\Big|\boldsymbol{Y}_n\right) \\
= & \frac{1}{D}\int_\sigma\int_{\Theta_n}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma) \\
= & J_1+J_2+J_3,
\end{aligned}
\tag{10}
$$

where
$J_1 = \frac{1}{D}\int_{R_1}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)$,
$J_2 = \frac{1}{D}\int_{R_2}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)$,
$J_3 = \frac{1}{D}\int_{R_3}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)$,
$R_1 = \{\theta_i \in [-a-c,a+c], \sigma \le \sigma_n\}$,
$R_2 = \{\theta_i \in [-a-c,a+c]^c, \sigma \le \sigma_n\}$,
$R_3 = \{\sigma > \sigma_n\}$.

Then it follows from Lemma 4.1 of Mukhopadhyay and Bhattacharya (2012a) that

$$J_3 \le H_1 P\left(R_3\big|\boldsymbol{Y}_n\right) \le H_1\epsilon_n^*, \tag{11}$$

where $H_1 = \sup_{\{y,\theta,\sigma\}}\left\{\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\right\} = \frac{1}{k}$.

Using Lemma 4.2 of Mukhopadhyay and Bhattacharya (2012a) we obtain

$$J_2 \le H_1 P\left(R_2\big|\boldsymbol{Y}_n\right) \le H_1 B_n. \tag{12}$$

$$J_1 = \frac{1}{(\upsilon_n(y)+k)}\phi\left(\frac{y-\mu_n^*(y)}{\upsilon_n(y)+k}\right)\left(1-P\left(R_2\big|\boldsymbol{Y}_n\right)-P\left(R_3\big|\boldsymbol{Y}_n\right)\right), \tag{13}$$

where, for every y, $\mu_n^*(y) \in (-a-c, a+c)$ and $\upsilon_n(y) \in (0,\sigma_n)$, applying $GMVT$.

Let us choose ϵ_n and σ_n in a way such that ϵ_n^* and B_n converge to zero as $n \to \infty$. Now we note that, in J_1, the range of θ_i remains same for all n. It is the range of σ that varies with n and the point $\upsilon_n(y)$ varies with n. This has the effect of varying $\mu_n^*(y)$ since the kernel depends upon both θ_i and σ. This implies that $\mu_n^*(y)$ and $\upsilon_n(y)$ depend upon σ_n, so that $\mu_n^*(y) = \mu(\sigma_n, y)$, $\upsilon_n(y) = \Psi(\sigma_n, y)$, such that

$\Psi(0,y)=0$ (note that $\upsilon_n(y)<\sigma_n$ and $\sigma_n\to 0$). We also assume that $\mu(0,y)=\mu^*(y)$. As before, we assume that μ and Ψ are continuously differentiable at least once; indeed we can choose μ and Ψ to be smooth functions such that $\mu_n^*(y)=\mu(\sigma_n,y)$, $\mu(0,y)=\mu^*(y)$, $\upsilon_n(y)=\Psi(\sigma_n,y)$, and $\Psi(0,y)=0$.

Then the following Taylor series expansion is valid (letting $x=\sigma_n$),

$$\begin{aligned}&\frac{1}{(\Psi(x,y)+k)}\phi\left(\frac{y-\mu(x,y)}{\Psi(x,y)+k}\right)\\ =\ &\frac{1}{k}\phi\left(\frac{y-\mu(0,y)}{k}\right)+x\frac{d}{dx}\left(\frac{1}{(\Psi(x,y)+k)}\phi\left(\frac{y-\mu(x,y)}{\Psi(x,y)+k}\right)\right)\Bigg|_{x=x^*},\end{aligned}$$

where x^* lies between 0 and x. Noting that the terms are bounded for any $y\in\mathbb{R}$, where $\mathbb{R}$ denotes the real line, we have,

$$\sup_{-\infty<y<\infty}\left|\frac{1}{(\upsilon_n(y)+k)}\phi\left(\frac{y-\mu_n^*(y)}{\upsilon_n(y)+k}\right)-\frac{1}{k}\phi\left(\frac{y-\mu(0,y)}{k}\right)\right|=O(\sigma_n). \tag{14}$$

Thus, for any y, we have that $\frac{1}{(\upsilon_n(y)+k)}\phi\left(\frac{y-\mu_n^*(y)}{\upsilon_n(y)+k}\right)\to\frac{1}{k}\phi\left(\frac{y-\mu^*(y)}{k}\right)$.

It follows from (14) that

$$\sup_{-\infty<y<\infty}\left|J_1-\frac{1}{k}\phi\left(\frac{y-\mu^*(y)}{k}\right)\right|=O\left(B_n+\epsilon_n^*+\sigma_n\right). \tag{15}$$

Finally, since B_n and ϵ_n^* can be made to converge to 0, $J_2,J_3\to 0$, $J_1\to\frac{1}{k}\phi\left(\frac{y-\mu^*(y)}{k}\right)$. Also it holds that

$$\begin{aligned}\int_y J_1dy &= \frac{1}{D}\int_y\int_{R_1}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)dy\\ &= 1-P\left(R_2\middle|\boldsymbol{Y}_n\right)-P\left(R_3\middle|\boldsymbol{Y}_n\right)\\ &\to 1. \qquad (16)\end{aligned}$$

Again, since we have proved above that $J_1\to\frac{1}{k}\phi\left(\frac{y-\mu^*(y)}{k}\right)$, if we can show that for all n, J_1 given by (13) is bounded above by an integrable function $h(y)$ for every y, then it follows from the dominated convergence theorem (DCT) that

$$\int_y J_1dy\to\int_y\frac{1}{k}\phi\left(\frac{y-\mu^*(y)}{k}\right)dy. \tag{17}$$

To show that J_1 is bounded above by an integrable function h, first note that $J_1\leq\frac{1}{k}\phi\left(\frac{y-\mu_n^*(y)}{\upsilon_n(y)+k}\right)$. We then define h to be the following.

$$\begin{aligned}h(y) &= \frac{1}{k}\ \text{ if }\ y\in[-\ell_1,\ell_2];\\ &= \frac{1}{k}\phi\left(\frac{y}{C^*}\right)\ \ \text{otherwise},\end{aligned}$$

where $-\ell_1 \ll -a-c$ and $\ell_2 \gg a+c$, and C^* satisfies

$$(C^*)^2 > \sup_{\mathcal{S}} \frac{y^2(\upsilon_n(y)+k)^2}{(y-\mu_n^*(y))^2},$$

where $\mathcal{S} = \{n, y \in [-\ell_1, \ell_2]^c, \mu_n^*(y) \in (-a-c, a+c), \upsilon_n(y) \in (0, \sigma_n), \sigma_n < K^*\}$. The right hand side of the above inequality is clearly finite. Moreover, it is easy to see that $J_1 \leq \frac{1}{k}\phi\left(\frac{y-\mu_n^*(y)}{\upsilon_n(y)+k}\right) \leq h(y)$ for every n, and almost all y, and that h is an integrable function. Hence, DCT holds.

It follows from (16) and (17) that

$$\int_y \frac{1}{k}\phi\left(\frac{y-\mu^*(y)}{k}\right) dy = 1, \tag{18}$$

showing that $f_0(y) = \frac{1}{k}\phi\left(\frac{y-\mu^*(y)}{k}\right)$ is a density. Indeed, we consider f_0 as the true data-generating density.

Now note that,

$$E\left(\hat{f}_{EW}(y \mid \Theta_n, \sigma)\middle|\boldsymbol{Y}_n\right) = O\left(\frac{\alpha}{\alpha+n}\right) + \frac{1}{\alpha+n}\sum_{i=1}^{n} E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right),$$

where the first term is thanks to (9). Further note that,

$$\frac{1}{\alpha+n}\sum_{i=1}^{n} E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right) = \frac{n}{\alpha+n} \times (J_1+J_2+J_3). \tag{19}$$

$$\begin{aligned}
&E\left(\hat{f}_{EW}(y \mid \Theta_n, \sigma)\middle|\boldsymbol{Y}_n\right) - f_0(y)\\
&= O\left(\frac{\alpha}{\alpha+n}\right) - \frac{\alpha}{\alpha+n} f_0(y)\\
&\quad + \frac{n}{\alpha+n}(J_1 - f_0(y)) + \frac{n}{\alpha+n}(J_2+J_3)
\end{aligned}$$

Since $f_0(y) \leq \frac{1}{k}$, we have

$$\begin{aligned}
&\left|E\left(\hat{f}_{EW}(y \mid \Theta_n, \sigma)\middle|\boldsymbol{Y}_n\right) - f_0(y)\right|\\
&\leq O\left(\frac{\alpha}{\alpha+n}\right) + \frac{n}{\alpha+n}\left(|J_2|+|J_3|\right)\\
&\quad + \frac{n}{\alpha+n}|J_1 - f_0(y)|
\end{aligned}$$

It follows from (11), (12), and (15) and the fact that the orders of J_1, J_2, J_3 are independent of y, that, for any $\boldsymbol{Y}_n \in \boldsymbol{S}_n$,

$$\sup_{-\infty<y<\infty}\left|E\left(\hat{f}_{EW}(y \mid \Theta_n,\sigma)\middle|\boldsymbol{Y}_n\right)-f_0(y)\right| = O\left(\frac{\alpha}{\alpha+n}+\frac{n}{\alpha+n}(B_n+\epsilon_n^*+\sigma_n)\right),$$

proving the theorem.

□

S-3. Proofs of results associated with Section 7 of MB.

S-3.1. *Proof of Lemma 7.2.*

PROOF. $P\left(\sigma>\sigma_n|\boldsymbol{Y}_n\right)=\frac{\sum_z\int_{\sigma_n}^{\infty}\int_{\Theta_{M_n}}L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)}{\sum_z\int_{0}^{\infty}\int_{\Theta_{M_n}}L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)}=\frac{N}{D}$, where $H(\Theta_{M_n})$ is the joint distribution of Θ_{M_n} and

$$L(\Theta_{M_n},z,\boldsymbol{Y}_n)=\prod_{j=1}^{M_n}\frac{1}{\sigma^{n_j}}e^{-\frac{1}{2}\sum_{t:z_t=j}\left(\frac{Y_t-\theta_j}{\sigma}\right)^2},$$

where $n_j=\#\{t: z_t=j\}$ (for any set $\mathbb{A}$, $\#\mathbb{A}$ denotes the cardinality of the set $\mathbb{A}$). Let $E^*=$\{all $\theta_l\in\Theta_{M_n}$ are in $[-c_1,c_1]$\}.

Then,

$$\begin{aligned}&\int_0^{\infty}\int_{\Theta_{M_n}}L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)\\ \geq\ &\int_{\sigma\in(b_n,\sigma_n)}\int_{\Theta_{M_n}\in E^*}L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma).\end{aligned} \tag{20}$$

Now note that

$$\begin{aligned}&|Y_t|<a, |\theta_j^*|<c_1\Rightarrow(Y_t-\theta_j^*)^2<(a+c_1)^2,\\ \Rightarrow\ &\sum_{j=1}^{M_n}\sum_{t:z_t=j}\left(Y_t-\theta_j^*\right)^2<n(a+c_1)^2.\end{aligned}$$

Again, from the Polya urn scheme we have $H(\Theta_{M_n})\geq\prod_{j=1}^{M_n}\frac{\alpha}{\alpha+M_n}G_0(\theta_j)$ which implies $P\left(\Theta_{M_n}\in E^*\right)\geq\left(\frac{\alpha}{\alpha+M_n}\right)^{M_n}H_0^{M_n}$, where $H_0=\int_{-c_1}^{c_1}G_0(\theta)d\theta$.

Hence, for $\sigma_n < \sqrt{n}(a + c_1)$,

$$D \geq M_n^n \frac{\exp\left(\frac{-n(a+c_1)^2}{2(b_n)^2}\right)}{b_n} P(b_n < \sigma \leq \sigma_n) \left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} H_0^{M_n}. \tag{21}$$

Again, in the same way as Lemma 6.2,$N \leq M_n^n \frac{1}{(\sigma_n)^n} P(\sigma > \sigma_n)$. Since $P(\sigma > \sigma_n) = O(\epsilon_n)$ and $P(b_n < \sigma \leq \sigma_n) = O(1 - \epsilon_n)$. Thus,

$$P(\sigma > \sigma_n | \boldsymbol{Y}_n) = O(\epsilon^*_{M_n}),$$

where

$$\epsilon^*_{M_n} = \frac{\epsilon_n}{1 - \epsilon_n} \exp\left(\frac{n(a + c_1)^2}{2(b_n)^2}\right) \frac{(\alpha + M_n)^{M_n}}{\alpha^{M_n} H_0^{M_n}}.$$

Hence, the proof follows. □

S-3.2. *Proof of Lemma 7.3.*

PROOF. Clearly, E^c ={at least one θ_k in likelihood is in $[-a - c, a + c]^c$}. We have

$$\begin{aligned} & P(Z \in R_1^*, \Theta_{M_n} \in E^c, \sigma \leq \sigma_n | \boldsymbol{Y}_n) \\ = \quad & \frac{\sum_{j=1}^{(M_n-1)} \sum_{l=1}^{j} \sum_{z \in V_j} \int_{\sigma \leq \sigma_n} \int_{W_l} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})}{\sum_z \int_\sigma \int_{\Theta_{M_n}} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})} \\ \leq \quad & \frac{\sum_{j=1}^{(M_n-1)} \sum_{l=1}^{j} \sum_{z \in V_j} \int_{\sigma \leq \sigma_n} \int_{W_l} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})}{\sum_{j=1}^{(M_n-1)} \sum_{z \in V_j} \int_\sigma \int_{\Theta_{M_n}} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})} \end{aligned} \tag{22}$$

Note that

$$L(\Theta_{M_n}, z, \boldsymbol{Y}_n) = \prod_{j=1}^{M_n} \frac{1}{\sigma^{n_j}} e^{-\frac{1}{2}\sum_{t:z_t=j}\left(\frac{Y_t - \theta_j}{\sigma}\right)^2} = \prod_{j=1}^{M_n} \frac{1}{\sigma^{n_j}} e^{-\frac{1}{2}\sum_{t:z_t=j}\left(\frac{Y_t - \bar{Y}_j}{\sigma}\right)^2} e^{-\frac{n_j}{2}\left(\frac{\bar{Y}_j - \theta_j}{\sigma}\right)^2}, \tag{23}$$

where $n_j = \#\{l : z_l = j\}$ and $\bar{Y}_j = \frac{\sum_{l:z_l=j} Y_l}{n_j}$.

Let $H_j(\theta_j \mid \Theta_{-jM_n})$ be the conditional distribution of θ_j given Θ_{-jM_n} and $H_{-j}(\Theta_{-jM_n})$ the joint distribution of Θ_{-jM_n}, where $\Theta_{-jM_n} = \Theta_{M_n} \setminus \theta_j$. Since

$$H_j(\theta_j \mid \Theta_{-jM_n}) = \frac{\alpha}{\alpha + M_n - 1} G_0(\theta_j) + \frac{1}{\alpha + M_n - 1} \sum_{l=1, l \neq j}^{M_n} \delta_{\theta_l}. \tag{24}$$

and $\sigma < a$, we have in the denominator for each $z \in V_j$,

$$\begin{aligned} & \int_{\theta_j} e^{-\frac{n_j}{2}\left(\frac{\bar{Y}_j-\theta_j}{\sigma}\right)^2} dH_j(\theta_j \mid \Theta_{-jM_n}) \\ \geq \; & \frac{\alpha}{\alpha+M_n}\int_{\bar{Y}_j-\frac{\sigma}{n_j^{1/2}}}^{\bar{Y}_j+\frac{\sigma}{n_j^{1/2}}} \exp\left(-\frac{n_j(\bar{Y}_j-\theta_j)^2}{2\sigma^2}\right) dG_0(\theta_j) \\ \geq \; & \frac{\alpha}{\alpha+M_n} e^{-1/2}\frac{\sigma}{n_j^{1/2}}\delta, \end{aligned} \tag{25}$$

where δ is the lower bound of the density of G_0 on $[\bar{Y}_j - \frac{\sigma}{n_j^{1/2}}, \bar{Y}_j + \frac{\sigma}{n_j^{1/2}}]$ (we assume that the density of G_0 is strictly positive in neighborhoods of $\bar{Y}_j$, for each j.

Thus for each $z \in V_j$ we have,

$$\begin{aligned} & \int_\sigma \int_{\Theta_{M_n}} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}) \\ \geq \; & \frac{\alpha}{(\alpha+M_n)n_j^{1/2}} e^{-1/2}\delta \\ & \times \int_0^a \int_{\Theta_{-jM_n}} \left[\frac{1}{\sigma^{(n-1)}} \prod_{l=1}^{M_n} e^{-\frac{1}{2}\sum_{t:z_t=l}\left(\frac{Y_t-\bar{Y}_l}{\sigma}\right)^2} \right. \\ & \left. \prod_{l=1,l\neq j}^{M_n} e^{-\frac{n_l}{2}\left(\frac{\bar{Y}_l-\theta_l}{\sigma}\right)^2} dH_{-j}(\Theta_{-jM_n}) dG_n(\sigma)\right]. \\ = \; & \frac{\alpha}{(\alpha+M_n)n_j^{1/2}} e^{-1/2}\delta \times \zeta_n(j,z), \quad \text{(say)}, \end{aligned} \tag{26}$$

where

$$\begin{aligned} \zeta_n(j,z) = \int_0^a \int_{\Theta_{-jM_n}} & \left[\frac{1}{\sigma^{(n-1)}} \prod_{l=1}^{M_n} e^{-\frac{1}{2}\sum_{t:z_t=l}\left(\frac{Y_t-\bar{Y}_l}{\sigma}\right)^2} \right. \\ & \left. \prod_{l=1,l\neq j}^{M_n} e^{-\frac{n_l}{2}\left(\frac{\bar{Y}_l-\theta_l}{\sigma}\right)^2} dH_{-j}(\Theta_{-jM_n}) dG_n(\sigma)\right]. \end{aligned} \tag{27}$$

To obtain a lower bound for the numerator we note that for each $z \in V_j$ and $j = 1(1)M_n$, $|\bar{Y}_j| < a$ (since each $|Y_l| < a, l = 1, \ldots, n$) and $\theta_j \in [-a-c, a+c]^c$.

This implies

$$\begin{aligned}
&\frac{1}{\sigma}\exp\left(-n_j(\bar{Y}_j-\theta_j)^2/2\sigma^2\right) \\
\leq\ & \frac{1}{n_j^{1/2}}\frac{n_j^{1/2}}{\sigma}\exp\left(-n_jc^2/4\sigma^2\right)\exp(-c^2/4\sigma_n^2) \\
\leq\ & \frac{A_1^*}{n_j^{1/2}}\exp(-c^2/4\sigma_n^2),
\end{aligned}$$

where $A_1^*=\sup_{\sigma,n_j}\left\{\frac{n_j^{1/2}}{\sigma}\exp\left(-\frac{n_jc^2}{4\sigma^2}\right)\right\}$. It is easy to check that A_1^* is free of n. Thus for each $z\in V_j$,

$$\begin{aligned}
&\int_{\sigma\leq\sigma_n}\int_{W_l}L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n}) \\
\leq\ & \frac{A_1^*}{n_j^{1/2}}\exp(-c^2/4\sigma_n^2) \\
&\times\int_0^{\sigma_n}\int_{\Theta_{-jM_n}}\left[\frac{1}{\sigma^{(n-1)}}\prod_{l=1}^{M_n}e^{-\frac{1}{2}\sum_{t:z_t=l}\left(\frac{Y_t-\bar{Y}_l}{\sigma}\right)^2}\right. \\
&\qquad\left.\times\prod_{l\neq j}e^{-\frac{n_l}{2}\left(\frac{\bar{Y}_l-\theta_l}{\sigma}\right)^2}dH_{-j}(\Theta_{-jM_n})dG_n(\sigma)\right] \\
\leq\ & A_1^*\exp(-c^2/4\sigma_n^2)\times\zeta_n(j,z).
\end{aligned}$$

As a result, using (22), we see that

$$\begin{aligned}
&P(Z\in R_1^*,\Theta_{M_n}\in E^c,\sigma\leq\sigma_n|\boldsymbol{Y}_n) \\
\leq\ & \frac{\sum_{j=1}^{(M_n-1)}\sum_{l=1}^{j}\sum_{z\in V_j}\int_{\sigma\leq\sigma_n}\int_{W_l}L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})}{\sum_{j=1}^{(M_n-1)}\sum_{z\in V_j}\int_\sigma\int_{\Theta_{M_n}}L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})} \\
\leq\ & \frac{A_1^*\exp(-c^2/4\sigma_n^2)\times\sum_{j=1}^{(M_n-1)}\sum_{l=1}^{j}\sum_{z\in V_j}\zeta_n(j,z)}{\frac{\alpha}{\alpha+M_n}e^{-1/2}\delta\times\sum_{j=1}^{(M_n-1)}\sum_{z\in V_j}\zeta_n(j,z)} \\
\leq\ & \frac{A_1^*\exp(-c^2/4\sigma_n^2)\times\sum_{j=1}^{(M_n-1)}\sum_{l=1}^{(M_n-1)}\sum_{z\in V_j}\zeta_n(j,z)}{\frac{\alpha}{\alpha+M_n}e^{-1/2}\delta\times\sum_{j=1}^{(M_n-1)}\sum_{z\in V_j}\zeta_n(j,z)} \\
=\ & \frac{(M_n-1)A_1^*\exp(-c^2/4\sigma_n^2)}{\frac{\alpha}{\alpha+M_n}e^{-1/2}\delta}, \qquad (28)
\end{aligned}$$

proving the lemma.

□

S-3.3. *Proof of Lemma 7.4.*

PROOF. Note that,

$$
\begin{aligned}
& P(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n | \boldsymbol{Y}_n) \\
= \; & \frac{\sum_{z \in R_1^*} \int_{\sigma \leq \sigma_n} \int_{\Theta_{M_n} \in E} \int L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})}{\sum_z \int_\sigma \int_{\Theta_{M_n}} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})} \\
= \; & \frac{N}{D}. \qquad (29)
\end{aligned}
$$

Let $\Theta_{-z} = \Theta_{M_n} \setminus \Theta_z$. Let us assume, without loss generality, that $\{\theta_1, \ldots, \theta_d\}$ is the set of θ_l's present in the likelihood with d being the number of such θ_l's for $z \in R_1^*$.

Since

$$
\begin{aligned}
L(\Theta_{M_n}, z, \boldsymbol{Y}_n) &= \frac{1}{\sigma^n} e^{-\frac{\sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2}{2\sigma^2}} \times e^{-\frac{\sum_{j=1}^{M_n} n_j (\bar{Y}_j - \theta_j)^2}{2\sigma^2}} \\
&\leq \frac{1}{\sigma^n} e^{-\frac{\sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2}{2\sigma^2}},
\end{aligned}
$$

it follows that

$$
\begin{aligned}
& \int_{\Theta_{M_n} \in E} \int_0^{\sigma_n} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dH(\Theta_{M_n}) dG_n(\sigma) \\
& \leq \int_{\theta_1 \in [-a-c, a+c]} \int_{\Theta_{-1M_n}} \int_0^{\sigma_n} \frac{1}{\sigma^n} e^{\frac{-\sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2}{2\sigma^2}} dH(\Theta_{M_n}) dG_n(\sigma) \\
& \leq A_1^* \int_{\theta_1 \in [-a-c, a+c]} \int_{\Theta_{-1M_n}} \int_0^{\sigma_n} dH(\Theta_{M_n}) dG_n(\sigma) \\
& = A_1^* G_0([-a-c, a+c]) O(1 - \epsilon_n), \qquad (30)
\end{aligned}
$$

where $A_1^* = \sup_{\{\sigma \in (0, \sigma_n)\}} \left\{ \frac{1}{\sigma^n} e^{-\frac{C_n^{(1)}}{2\sigma^2}} \right\} = \left(\frac{1}{\sigma_n} \right)^n e^{-\frac{C_n^{(1)}}{2\sigma_n^2}}$,
$C_n^{(1)} = \inf_{\{z \in R_1^*\}} \left(\sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2 \right)$.

Clearly, for each $z \in R_1^*$ each term in N is bounded above by

$$
N^* = \left(\frac{1}{\sigma_n} \right)^n e^{-\frac{C_n^{(1)}}{2\sigma_n^2}} \times G_0([-a-c, a+c]) \times O(1 - \epsilon_n).
$$

Hence

$$N \leq (M_n - 1)^n \, N^* \tag{31}$$

Let $C_n^{(2)} = \sup_{z \in R_1^*} \sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2$. Now, assuming that k_n is a sequence diverging to ∞ and denoting R^*=$\{\theta_1 \in [\bar{Y}_1 - k_n, \bar{Y}_1 + k_n], \ldots, \theta_d \in [\bar{Y}_d - k_n, \bar{Y}_d + k_n]$, rest θ_l's are in $(-\infty, \infty)$, $nk_n \leq \sigma \leq 2nk_n\}$,

$$\begin{aligned}
D &= \int_{\Theta_{M_n}} \int_0^\infty \frac{1}{\sigma^n} \left[e^{-\frac{\sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2}{2\sigma^2}} \right. \\
&\qquad \left. \times \, e^{-\frac{\sum_{j=1}^{M_n} n_j (\bar{Y}_j - \theta_j)^2}{2\sigma^2}} dH(\Theta_{M_n}) dG_n(\sigma) \right] \\
&\geq \int_{R^*} \left[\frac{1}{\sigma^n} e^{-\frac{\sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2}{2\sigma^2}} \right. \\
&\qquad \left. \times \, e^{-\frac{\sum_{j=1}^{M_n} n_j (\bar{Y}_j - \theta_j)^2}{2\sigma^2}} dH(\Theta_{M_n}) dG_n(\sigma) \right] \\
&\geq \inf_{\{z, \sigma \in [nk_n, 2nk_n]\}} \left(\frac{1}{\sigma^n} e^{-\frac{\sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2}{2\sigma^2}} \right) \\
&\quad \times \int_{R^*} e^{\frac{\sum_{j=1}^{d} n_j (\bar{Y}_j - \theta_j)^2}{2\sigma^2}} \times dH(\Theta_{M_n}) dG_n(\sigma) \\
&\geq \left(\frac{1}{2nk_n} \right)^n e^{-\frac{C_n^{(2)}}{8n^2 k_n^2}} \times \int_{R^*} e^{-\frac{nk_n^2}{2\sigma^2}} \times dH(\Theta_{M_n}) dG_n(\sigma) \\
&\geq \left(\frac{1}{2nk_n} \right)^n e^{-\frac{C_n^{(2)}}{8n^2 k_n^2}} \times e^{-\frac{1}{2n}} \times \int_{R^*} dH(\Theta_{M_n}) dG_n(\sigma),
\end{aligned}$$

because $\sigma \geq nk_n \Rightarrow e^{-\frac{nk_n^2}{2\sigma^2}} \geq e^{-\frac{1}{2n}}$. Thus,

$$
\begin{aligned}
D \quad &\geq \quad M_n^n \left(\frac{1}{2nk_n}\right)^n e^{-\frac{C_n^{(2)}}{8n^2k_n^2}} \times \prod_{j=1}^{d} G_0([\bar{Y}_j - k_n, \bar{Y}_j + k_n]) \times e^{-\frac{1}{2n}} \\
&\quad \times \left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} \times O(\epsilon_n), \qquad (32)
\end{aligned}
$$

assuming that $\int_{nk_n}^{2nk_n} dG_n(\sigma) = O(\epsilon_n)$ as well.

Inequalities (31) and (32) imply that $\frac{N}{D}$ is of the order

$$
\frac{(M_n - 1)^n \left(\frac{1}{\sigma_n}\right)^n e^{-\frac{C_n^{(1)}}{2\sigma_n^2}} \times G_0([-a-c, a+c]) \times O(1-\epsilon_n)}{M_n^n \left(\frac{1}{2nk_n}\right)^n e^{-\frac{C_n^{(2)}}{8n^2k_n^2}} e^{-\frac{1}{2n}} \left(\frac{\alpha}{\alpha+M_n}\right)^{M_n} \prod_{j=1}^{d} G_0([\bar{Y}_j - k_n, \bar{Y}_j + k_n]) \times O(\epsilon_n)}, \qquad (33)
$$

where $\prod_{j=1}^{d} G_0([\bar{Y}_j - k_n, \bar{Y}_j + k_n]) \to 1$ as $n \to \infty$.

As shown in Section 2 of MB, $\sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2$ converges to σ_T^2 almost surely. Thus for large n the possible values of $\sum_{j=1}^{M_n} \sum_{t:z_t=j} (Y_t - \bar{Y}_j)^2$ will be close to σ_T^2 almost surely. From this we can say that for large n, it holds, almost surely, that $C_n^{(1)} \sim C_n^{(2)} \sim C_n$. We now investigate the appropriate order of C_n such that

$$
\left(\frac{1}{b_n\sigma_n}\right)^n e^{-\frac{C_n}{2b_n^2\sigma_n^2}} G_0([-a-c, a+c]) O(1-\epsilon_n) \lesssim \left(\frac{1}{2nk_n}\right)^n e^{-\frac{C_n}{8n^2k_n^2}} O(\epsilon_n) \qquad (34)
$$

holds for large n.

Taking logarithm of both sides of (34) yields

$$
\begin{aligned}
& n\log\left(\frac{1}{b_n\sigma_n}\right)-\frac{C_n}{2b_n^2\sigma_n^2}+O\left(\log\left(\frac{1-\epsilon_n}{\epsilon_n}\right)\right)+\log\left(H_0\right)\\
\lesssim\ & n\log\left(\frac{1}{2nk_n}\right)-\frac{C_n}{8n^2k_n^2}\\
\Leftrightarrow\ & C_n\left(\frac{1}{2b_n^2\sigma_n^2}-\frac{1}{8n^2k_n^2}\right)\gtrsim n\log\left(\frac{1}{b_n\sigma_n}\right)-n\log\left(\frac{1}{2nk_n}\right)\\
& \qquad +O\left(\log\left(\frac{1-\epsilon_n}{\epsilon_n}\right)\right)+\log\left(H_0\right)\\
\Leftrightarrow\ & C_n\gtrsim\frac{n\left[\log\left(\frac{1}{2nk_n}\right)+\log(2nk_n)\right]+O\left(\log\left(\frac{1-\epsilon_n}{\epsilon_n}\right)\right)}{\left(\frac{1}{2b_n^2\sigma_n^2}-\frac{1}{8n^2k_n^2}\right)}\\
& \qquad +\frac{\log\left(H_0\right)}{\left(\frac{1}{2b_n^2\sigma_n^2}-\frac{1}{8n^2k_n^2}\right)}\\
\Leftrightarrow\ & C_n\gtrsim\frac{n\left[\log\left(\frac{1}{b_n\sigma_n}\right)+O\left(\frac{1}{n}\log\left(\frac{1-\epsilon_n}{\epsilon_n}\right)\right)\right]}{\left(\frac{1}{b_n^2\sigma_n^2}\right)}. \qquad (35)
\end{aligned}
$$

Thus, (34) holds for C_n given by (17) of MB. Hence, (18) of MB holds under the additional assumption (17) of MB.

□

S-3.4. *Proof of Theorem 7.1.*

PROOF.

$$
\begin{aligned}
& E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right)\\
=\ & \frac{\sum_z\int_{\Theta_{M_n}}\int_\sigma\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)}{\sum_z\int_{\Theta_{M_n}}\int_\sigma L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)}\\
=\ & \frac{N}{D}, \qquad (36)
\end{aligned}
$$

where $L(\Theta_{M_n},z,\boldsymbol{Y}_n)=\prod_{j=1}^{M_n}e^{-\frac{\sum_{t:z_t=j}(Y_t-\bar{Y}_j)^2}{2\sigma^2}}e^{-\frac{n_j(\bar{Y}_j-\theta_j)^2}{2\sigma^2}}$, likelihood of Θ_{M_n}, $n_j=\#\{k:z_k=j\}$, $\bar{Y}_j=\frac{1}{n_j}\sum_{t:z_t=j}Y_t$.

To simplify the calculations we can split the set of all of z's in to R_1^* and $(R_1^*)^c$; the cardinality of the set of all z-vectors satisfying these conditions are $(M_n-1)^n$

and $M_n^n - (M_n - 1)^n$, respectively. Denote $I_1 = \{\Theta_{M_n} \in E^c, \sigma \leq \sigma_n\}$, $I_2 = \{\Theta_{M_n} \in E, \sigma \leq \sigma_n\}$, $I_3 = \{\theta_i \in [-a-c, a+c]^c, \sigma \leq \sigma_n\}$, $I_4 = \{\theta_i \in [-a-c, a+c], \sigma \leq \sigma_n\}$, $I_5 = \{\sigma > \sigma_n\}$, where E has been defined in Section 7 of MB. Note that

$$\begin{aligned}
\{R_1^* \cap I_1\} \cup \{R_1^* \cap I_2\} &= R_1^* \cap \{\sigma \leq \sigma_n\} \\
\{(R_1^*)^c \cap I_3\} \cup \{(R_1^*)^c \cap I_4\} &= (R_1^*)^c \cap \{\sigma \leq \sigma_n\} \\
(\{R_1^* \cap I_1\} \cup \{R_1^* \cap I_2\}) \cup (\{(R_1^*)^c \cap I_3\} \cup \{(R_1^*)^c \cap I_4\}) &= \{\sigma \leq \sigma_n\}
\end{aligned}$$

We write

$$E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right) = S_1 + S_2 + S_3 + S_4 + S_5 \tag{37}$$

where,
$S_1 = \frac{1}{D}\sum_{R_1^*}\int_{I_1}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$,
$S_2 = \frac{1}{D}\sum_{R_1^*}\int_{I_2}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$,
$S_3 = \frac{1}{D}\sum_{(R_1^*)^c}\int_{I_3}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$,
$S_4 = \frac{1}{D}\sum_{(R_1^*)^c}\int_{I_4}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$,
$S_5 = \frac{1}{D}\sum_{z}\int_{I_5}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$.

Also let
$P_1 = P(Z \in R_1^*, \Theta_{M_n} \in E^c, \sigma \leq \sigma_n|\boldsymbol{Y}_n)$,
$P_2 = P(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n|\boldsymbol{Y}_n)$,
$P_3 = P(Z \in (R_1^*)^c, \theta_i \in [-a-c, a+c]^c, \sigma \leq \sigma_n|\boldsymbol{Y}_n)$,
$P_4 = P(Z \in (R_1^*)^c, \theta_i \in [-a-c, a+c], \sigma \leq \sigma_n|\boldsymbol{Y}_n)$,
$P_5 = P(\sigma \geq \sigma_n|\boldsymbol{Y}_n)$.

Let $H_1^* = \sup_{\{y,\theta_j,\sigma\}}\left\{\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)\right\} = \frac{1}{k}$. Then the upper bounds of the terms $S_1, \ldots, S_5$ are given as follows.

$$S_1 \leq H_1^* P(Z \in R_1^*, \Theta_{M_n} \in E^c, \sigma \leq \sigma_n|\boldsymbol{Y}_n) \leq H_1^*(M_n - 1)B_{M_n}, \tag{38}$$

from Lemma 7.3 of MB.

(39)
$$S_2 \leq H_1^* P(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n|\boldsymbol{Y}_n) \leq H_1^*\left(1 - \frac{1}{M_n}\right)^n\left(\frac{\alpha + M_n}{\alpha}\right)^{M_n},$$

from Lemma 7.4 of MB.

$$S_3 \le H_1^* P(Z \in (R_1^*)^c, \theta_i \in [-a-c, a+c]^c, \sigma \le \sigma_n \mid \boldsymbol{Y}_n) \le H_1^* B_{M_n}, \tag{40}$$

from Lemma 7.5 of MB.

$$S_5 \le H_1^* P(Z \in (R_1^*)^c, \sigma > \sigma_n \mid \boldsymbol{Y}_n) \le H_1^* \epsilon^*_{M_n}, \tag{41}$$

from Lemma 7.2 of MB.

$$\begin{aligned}
S_4 &= \frac{1}{D}\int_{I_4} \frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\sum_{z\in(R_1^*)^c} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dH(\Theta_{M_n}) dG_n(\sigma)\\
&= \frac{1}{D}\frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y-\theta_n^*(y)}{\sigma_n^*(y)+k}\right)\\
&\quad \times \sum_{z\in(R_1^*)^c}\int\int_{\theta_i\in[-a-c,a+c]}\int_0^{\sigma_n} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dH(\Theta_{M_n}) dG_n(\sigma) \qquad (42)\\
&= \frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y-\theta_n^*(y)}{\sigma_n^*(y)+k}\right)\\
&\quad \times P(Z\in(R_1^*)^c, \theta_i \in [-a-c, a+c], \sigma \le \sigma_n \mid \boldsymbol{Y}_n)\\
&= \frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y-\theta_n^*(y)}{\sigma_n^*(y)+k}\right)(1-P_1-P_2-P_3-P_5),
\end{aligned}$$
(43)

where (42) is obtained by using $GMVT$, $\theta_n^*(y) \in (-a-c, a+c)$, and $\sigma_n^*(y) \in (0, \sigma_n)$.
The integration and summation can be interchanged since the number of terms under summation is finite for a particular value of n.

Note that equations (38)–(41), and P_1, P_2, P_3, P_5 converge to zero under proper conditions. In particular, P_1 converges to 0 if σ_n is chosen to be sufficiently small. Also b_n can also cosen to be very small such that it satisfies $b_n^2 < \sigma_n^2$ for all n.

These choices get P_3 to converge to 0 and P_5 converges to zero if $\frac{\epsilon_n}{1-\epsilon_n} \prec \left(\exp\left(\frac{n(2a+c)^2}{2(b_n)^2}\right)\frac{(\alpha+M_n)^{M_n}}{(\alpha)^{M_n}H_0^{M_n}}\right)^{-1}$.

P_2 converges to zero if $M_n \prec \sqrt{n}$, however, the form of the bound (18) given by Lemma 7.4 of MB is valid if (17) of MB holds.

Now note that $S_1+S_2+S_3+S_5 \leq H_1^*\left[P_1+P_2+P_3+P_5\right]$. Since under the specified assumptions P_1, P_2, P_3, P_5 converge to 0, as $n\rightarrow\infty$, the sum also goes to 0, as $n\rightarrow\infty$. Thus in S_4, the term $(1-P_1-P_2-P_3-P_5)\rightarrow 1$ as $n\rightarrow\infty$. Uniform convergence of $\frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y_1-\theta_n^*(y)}{\sigma_n^*(y)+k}\right)$ to $\frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)$ can be proved in exactly the same way using Taylor's series expansion as done in the case of the EW model. In particular, it holds that

$$\sup_{-\infty<y<\infty}\left|\frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y-\theta_n^*(y)}{\sigma_n^*(y)+k}\right)-\frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)\right|=O(\sigma_n), \tag{44}$$

We also conclude that for $\boldsymbol{Y}_n\in\boldsymbol{S}_n$,

$$\begin{aligned}
& E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\Big|\boldsymbol{Y}_n\right)-\frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)\\
= \; & S_1+S_2+S_3+S_5\\
& +\frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y_1-\theta_n^*(y)}{\sigma_n^*(y)+k}\right)(1-P_1-P_2-P_3-P_5)-\frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)\\
= \; & S_1+S_2+S_3+S_5-\frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y_1-\theta_n^*(y)}{\sigma_n^*(y)+k}\right)(P_1+P_2+P_3+P_5)\\
& +\frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y_1-\theta_n^*(y)}{\sigma_n^*(y)+k}\right)-\frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)\\
= \; & O\left((M_n-1)B_{M_n}\right)+O\left(\left(1-\frac{1}{M_n}\right)^{M_n}\left(\frac{\alpha+M_n}{\alpha}\right)^{M_n}\right)\\
& +O\left(B_{M_n}\right)+O\left(\epsilon^*_{M_n}\right)+O(\sigma_n)\\
= \; & O\left(M_nB_{M_n}+\left(1-\frac{1}{M_n}\right)^n\left(\frac{\alpha+M_n}{\alpha}\right)+\epsilon^*_{M_n}+\sigma_n\right).
\end{aligned} \tag{45}$$

Noting that the terms involving y are bounded above by $1/k$, it follows that

$$\begin{aligned}
& \sup_{-\infty<y<\infty}\left|E\left(\hat{f}_{SB}(y\mid\Theta_{M_n},\sigma)\mid\boldsymbol{Y}_n\right)-\frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)\right|\\
& \leq \sup_{-\infty<y<\infty}\left|E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\Big|\boldsymbol{Y}_n\right)-\frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)\right|\\
& = O\left(M_nB_{M_n}+\left(1-\frac{1}{M_n}\right)^n\left(\frac{\alpha+M_n}{\alpha}\right)+\epsilon^*_{M_n}+\sigma_n\right),
\end{aligned} \tag{46}$$

proving the theorem. □

S-3.5. *Proof of Theorem 7.6.*

PROOF. Recall that $J_1 = \frac{1}{D_1}\int_{I_4}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)$, and
$S_4 = \frac{1}{D_2}\int_{I_4}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\sum_{z\in(R_1^*)^c}L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$, where D_1 and D_2 denote the normalizing constants of the posteriors corresponding to the EW and the SB models, respectively.
Let $L=\max(M_n,n)$. Then,
$|J_1-S_4|$

$$
\begin{aligned}
&= \Bigg|\int_{I_4}\Bigg[\frac{1}{D_1}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\times L(\Theta_n,\boldsymbol{Y}_n)\\
&\qquad -\frac{1}{D_2}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\times\sum_{z\in(R_1^*)^c}L(\Theta_{M_n},z,\boldsymbol{Y}_n)\Bigg]dH(\Theta_L)dG_n(\sigma)\Bigg|\\
&= \frac{1}{(\sigma_{1,n}(y)+k)}\phi\left(\frac{y-\theta_{1,n}(y)}{\sigma_{1,n}(y)+k}\right)\times\left|P\left(I_4|\boldsymbol{Y}_n\right)-P\left((R_1^*)^c,I_4|\boldsymbol{Y}_n\right)\right| \qquad (47)\\
&\to 0. \qquad (48)
\end{aligned}
$$

Step (47) follows using $GMVT$, where the notation have the usual meanings, and step (48) follows because the first factor remains bounded and the second factor goes to zero (since $P\left(I_4|\boldsymbol{Y}_n\right)\to 1$, and $P\left((R_1^*)^c,I_4|\boldsymbol{Y}_n\right)\to 1$). In other words, J_1 and S_4 converge to the same model. Hence, we must have $\mu^*(y)=\theta^*(y)$. $\square$

S-4. Proof of results associated with Section 8 of MB.

S-4.1. *Proof of Lemma 8.2.*

PROOF. Note that

$$
\begin{aligned}
&Var\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\Bigg|\boldsymbol{Y}_n\right)\\
&= E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)-E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\Bigg|\boldsymbol{Y}_n\right)\Bigg|\boldsymbol{Y}_n\right)^2\\
&= J_1^{'}+J_2^{'}+J_3^{'}, \qquad (49)
\end{aligned}
$$

where
$J_1^{'}=\frac{1}{D}\int_{R_1}\xi_{in}^2\times L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)$,
$J_2^{'}=\frac{1}{D}\int_{R_2}\xi_{in}^2\times L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)$,

$J_3^{'}=\frac{1}{D}\int_{R_3}\xi^2_{in}\times L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)$,
$\xi_{in}=\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)-E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right)$.

Clearly,

$$J_2^{'}\leq H_3^2\times B_n, \tag{50}$$

and

$$J_3^{'}\leq H_3^2\times \epsilon^*_n, \tag{51}$$

where $H_3=\sup_{\{y,\theta_i,\sigma\}}|\xi_{in}|\leq\frac{2}{k}$. Abusing notation a bit, we re-define $H_3=\frac{2}{k}$.

Denoting $P_1=P\left(R_1\big|\boldsymbol{Y}_n\right),P_2=P\left(R_2\big|\boldsymbol{Y}_n\right),P_3=P\left(R_3\big|\boldsymbol{Y}_n\right)$, we concentrate on the term

$$\begin{aligned}J_1^{'}&=\frac{1}{D}\int\int_{\theta_i\in[-a-c,a+c]}\int_{\sigma\leq\sigma_n}\xi^2_{in}\times L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)\\&=\left[\frac{1}{(\tau_n(y)+k)}\phi\left(\frac{y-m_n(y)}{\tau_n(y)+k}\right)-E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right)\right]\\&\quad\times\frac{1}{D}\int\int_{\theta_i\in[-a-c,a+c]}\int_{\sigma\leq\sigma_n}\xi_{in}\times L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)\end{aligned} \tag{52}$$

applying $GMVT$, where, for every y, $m_n(y)\in(-a-c,a+c)$, and $\tau_n(y)\in(0,\sigma_n)$.

Now we consider the following term:

$$\begin{aligned}&\frac{1}{D}\int\int_{\theta_i\in[-a-c,a+c]}\int_{\sigma\leq\sigma_n}\xi_{in}\times L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)\\&=\frac{1}{D}\int\int_{\theta_i\in[-a-c,a+c]}\int_{\sigma\leq\sigma_n}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)\\&\quad-E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right)P\left(\theta_i\in[-a-c,a+c],\sigma\leq\sigma_n\bigg|\boldsymbol{Y}_n\right)\\&=J_1^{''}+J_2^{''}.\end{aligned} \tag{53}$$

For the part $J_1^{''}$, we note that

$$
\begin{aligned}
& \frac{1}{D}\int\int_{\theta_i\in[-a-c,a+c]}\int_{\sigma\leq\sigma_n}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)\\
= & \frac{1}{(\upsilon_n(y)+k)}\phi\left(\frac{y-\mu_n^*(y)}{\upsilon_n(y)+k}\right)(1-P_2-P_3).
\end{aligned}
\tag{54}
$$

From (13) and following Theorem 2.4 it follows that

$$
\begin{aligned}
& E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\Bigg|\boldsymbol{Y}_n\right)\\
= & J_2+J_3+\frac{1}{(\upsilon_n(y)+k)}\phi\left(\frac{y-\mu_n^*(y)}{\upsilon_n(y)+k}\right)(1-P_2-P_3).
\end{aligned}
\tag{55}
$$

Thus,

$$
\begin{aligned}
& J_1^{''}+J_2^{''}\\
= & \frac{1}{(\upsilon_n(y)+k)}\phi\left(\frac{y-\mu_n^*(y)}{\upsilon_n(y)+k}\right)\left[(1-P_2-P_3)-(1-P_2-P_3)^2\right]\\
& -(J_2+J_3)(1-P_2-P_3)\\
\leq & \frac{1}{(\upsilon_n(y)+k)}\phi\left(\frac{y-\mu_n^*(y)}{\upsilon_n(y)+k}\right)(P_2+P_3)(1-P_2-P_3)\\
& +H_1(P_2+P_3)(1-P_2-P_3)\\
= & O\left(B_n+\epsilon_n^*\right),
\end{aligned}
\tag{56}
$$

since $\frac{1}{(\upsilon_n(y)+k)}\phi\left(\frac{y-\mu_n^*(y)}{\upsilon_n+k}\right)\leq\frac{1}{k}$, $P_2=O(B_n)$ and $P_3=O(\epsilon_n^*)$.

□

S-4.2. *Proof of Lemma 8.3.*

PROOF. The covariance term can be written as,

$$
\begin{aligned}
& cov_{ij} = J_1^* + J_2^* + J_3^*, \quad where \\
& J_1^* = \frac{1}{D}\int_{R_1'} \xi_{in}\xi_{jn} L(\Theta_n, \boldsymbol{Y}_n) dH(\Theta_n) dG_n(\sigma), \\
& J_2^* = \frac{1}{D}\int_{R_2'} \xi_{in}\xi_{jn} L(\Theta_n, \boldsymbol{Y}_n) dH(\Theta_n) dG_n(\sigma), \\
& J_3^* = \frac{1}{D}\int_{R_3'} \xi_{in}\xi_{jn} L(\Theta_n, \boldsymbol{Y}_n) dH(\Theta_n) dG_n(\sigma), \\
& R_1^{'} = \{\theta_i \in [-a-c, a+c], \theta_j \in [-a-c, a+c], \sigma \leq \sigma_n\}, \\
& R_2^{'} = \{\text{at least one of } \theta_i \text{ and } \theta_j \text{ is in } [-a-c, a+c]^c, \sigma \leq \sigma_n\}, \\
& R_3^{'} = \{\sigma > \sigma_n\}. \qquad (57)
\end{aligned}
$$

Also denote $P_1^{'} = P\left(R_1^{'}\middle|\boldsymbol{Y}_n\right), P_2^{'} = P\left(R_2^{'}\middle|\boldsymbol{Y}_n\right), P_3^{'} = P\left(R_3^{'}\middle|\boldsymbol{Y}_n\right)$. Note that,

$$
J_3^* \leq H_3^2 P_3^{'} \leq H_1^2 \epsilon_n^*, \qquad (58)
$$

and

$$
\begin{aligned}
J_2^* &\leq H_3^2 P_2^{'} \\
&\leq H_3^2 P\left(\theta_i \in [a-c, a+c]^c, \sigma \leq \sigma_n \middle| \boldsymbol{Y}_n\right) \\
&\quad + H_3^2 P\left(\theta_j \in [a-c, a+c]^c, \sigma \leq \sigma_n \middle| \boldsymbol{Y}_n\right) \\
&\leq 2H_3^2 B_n. \qquad (59)
\end{aligned}
$$

Consider the term

$$
\begin{aligned}
J_1^* &= \frac{1}{D}\int_{R_1'} \xi_{in}\xi_{jn} L(\Theta_n, \boldsymbol{Y}_n) dH(\Theta_n) dG_n(\sigma) \\
&= \frac{1}{D}\xi_{in}^* \int_{R_1'} \xi_{jn} L(\Theta_n, \boldsymbol{Y}_n) dH(\Theta_n) dG_n(\sigma)
\end{aligned}
$$

(60)

where $\xi_{in}^* = \left[\frac{1}{(\eta_{1n}(y)+k)}\phi\left(\frac{y-\varphi_{1n}(y)}{\eta_{1n}(y)+k}\right) - E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right)\right], \varphi_{1n} \in (-a-c, a+c)$, and $\eta_{1n} \in (0, \sigma_n)$. Equation (60) is obtained by applying $GMVT$. We will study convergence of the term
$\int_{R_1'} \xi_{jn} L(\Theta_n, \boldsymbol{Y}_n) dH(\Theta_n) dG_n(\sigma)$.

$$
\begin{aligned}
& \frac{1}{D}\int_{R_1'} \xi_{jn} L(\Theta_n, \boldsymbol{Y}_n) dH(\Theta_n) dG_n(\sigma) \\
= & \quad \frac{1}{(\eta_{2n}(y)+k)}\phi\left(\frac{y-\varphi_{2n}(y)}{\eta_{2n}(y)+k}\right)(1-P_2^{'}-P_3^{'}) \\
- & \quad \left[\left(J_2+J_3+L_1^{''}+\frac{1}{(\eta_{2n}(y)+k)}\phi\left(\frac{y-\varphi_{2n}(y)}{\eta_{2n}(y)+k}\right)\left(1-P_2^{'}-P_3^{'}\right)\right)\times\right. \\
& \qquad \left.\left(1-P_2^{'}-P_3^{'}\right)\right],
\end{aligned}
$$

where

$$
\begin{aligned}
L_1^{''} &= \frac{1}{D}\int_{R_1^{''}} \frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right) L(\Theta_n, \boldsymbol{Y}_n) dH(\Theta_n) dG_n(\sigma) \\
&\leq H_1 P(R_1^{''}) \\
&\leq H_1 P(R_2^{'}) \\
&\leq H_1 B_n, \qquad (61)
\end{aligned}
$$

From (61), the equations (58), (59) (the latter two showing that $P_2^{'} = O(B_n)$, $P_3^{'} = O(\epsilon_n^*)$), and $\xi_{in}^* = O(1)$, it follows that $J_1^* = O\left(B_n+\epsilon_n^*\right)$. Finally, we have,

$$cov_{ij} = O\left(B_n+\epsilon_n^*\right)$$

□

S-5. Proofs of results associated with Section 9 of MB.

S-5.1. *Proof of Lemma 9.2.*

$$
\begin{aligned}
& Var\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right) \\
= & \quad E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)-E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right)\bigg|\boldsymbol{Y}_n\right)^2
\end{aligned}
$$

As in (45) we begin with splitting up the range of z and the range of integration of Θ_{M_n} and σ in the following way:

$$
\begin{aligned}
& E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)-E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right)\bigg|\boldsymbol{Y}_n\right)^2 \\
= & \quad S_1^*+S_2^*+S_3^*+S_4^*+S_5^*, \qquad (62)
\end{aligned}
$$

where S_i^* has same ranges of z, Θ_{M_n} and σ as S_i in Theorem 7.1 of MB only the integrand of the former is now replaced by
$\zeta_{M_n} = \left[\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) - E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right)\right]$, that is
$S_1^* = \frac{1}{D}\sum_{R_1^*}\int_{I_1}\zeta_{M_n}^2 L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$,
$S_2^* = \frac{1}{D}\sum_{R_1^*}\int_{I_2}\zeta_{M_n}^2 L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$,
$S_3^* = \frac{1}{D}\sum_{(R_1^*)^c}\int_{I_3}\zeta_{M_n}^2 L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$,
$S_4^* = \frac{1}{D}\sum_{(R_1^*)^c}\int_{I_4}\zeta_{M_n}^2 L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$,
$S_5^* = \frac{1}{D}\sum_{z}\int_{I_5}\zeta_{M_n}^2 L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)$.

Let $H_2^* = \sup_{\{y,\theta_i,\sigma\}}\left|\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) - E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right)\right| = \frac{2}{k}$. Then in the same way as in equations (38)–(41) it follows that

$$S_1^* \leq (H_2^*)^2 P\left(Z \in R_1^*, \Theta_{M_n} \in E^c, \sigma \leq \sigma_n \mid \boldsymbol{Y}_n\right) \leq (H_2^*)^2 (M_n - 1) B_{M_n}, \tag{63}$$

$$\begin{aligned} S_2^* &\leq (H_2^*)^2 P\left(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n \mid \boldsymbol{Y}_n\right) \\ &\leq (H_2^*)^2 \left(1 - \frac{1}{M_n}\right)^n \left(\frac{\alpha + M_n}{\alpha}\right)^{M_n}, \end{aligned} \tag{64}$$

$$S_3^* \leq (H_2^*)^2 P\left(Z \in (R_1^*)^c, \theta_i \in [-a-c, a+c]^c, \sigma \leq \sigma_n \mid \boldsymbol{Y}_n\right) \leq (H_2^*)^2 B_{M_n}, \tag{65}$$

$$S_5^* \leq (H_2^*)^2 P\left(Z \in (R_1^*)^c, \theta_i \in [-a-c, a+c], \sigma \leq \sigma_n \mid \boldsymbol{Y}_n\right) \leq (H_2^*)^2 \epsilon_{M_n}^*. \tag{66}$$

$$\begin{aligned} S_4^* &= \frac{1}{D}\int_{\Theta_{-iM_n}}\int_{\theta_i\in[-a-c,a+c]}\int_{\sigma\leq\sigma_n}\zeta_{M_n}^2 L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma) \\ &= \left[\frac{1}{\{(\sigma_n^v(y)+k)\}^2}\phi^2\left(\frac{y-\theta_n^v(y)}{\sigma_n^v(y)+k}\right) - E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right)\right] \times \\ &\quad \frac{1}{D}\int_{\Theta_{-iM_n}}\int_{\theta_i\in[-a-c,a+c]}\int_{\sigma\leq\sigma_n}\zeta_{M_n} L(\Theta_{M_n}, z, \boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma). \end{aligned} \tag{67}$$

Let $R^{'}$ =$\{\theta_i \in [-a-c, a+c]$, rest θ_l's are in $(-\infty,\infty)$, $\sigma \leq \sigma_n\}$. Then we consider the following:

$$\begin{aligned}
& \frac{1}{D}\int_{R'} \zeta_{M_n} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dH(\Theta_{M_n}) dG_n(\sigma) \\
= & \frac{1}{D}\int_{R'} \frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dH(\Theta_{M_n}) dG_n(\sigma) \\
& -E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right) P\left(\theta_i \in [-a-c, a+c], \sigma \leq \sigma_n | \boldsymbol{Y}_n\right) \\
= & S_1^{''} + S_2^{''}, \text{ say.}
\end{aligned} \tag{68}$$

The terms $S_1^{''}$ and $S_2^{''}$ can be dealt with in the same way as $J_1^{''}$ and $J_2^{''}$ were handled in the corresponding EW case and it can be shown that

$$S_4^* = O\left(MB_{M_n} + \left(1 - \frac{1}{M_n}\right)^n \left(\frac{\alpha + M_n}{\alpha}\right)^{M_n} + \epsilon^*_{M_n}\right). \tag{69}$$

Thus, $\sum_{i=1}^4 S_i^* = O\left(MB_{M_n} + \left(1 - \frac{1}{M_n}\right)^n \left(\frac{\alpha+M_n}{\alpha}\right)^{M_n} + \epsilon^*_{M_n}\right)$. Hence, the lemma follows.

S-5.2. *Proof of Lemma 9.3.* Let
$f_1(y,\theta_i,\theta_j,\sigma) = \left[\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) - E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right)\right] \times$
$\left[\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right) - E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)\middle|\boldsymbol{Y}_n\right)\right]$.
The covariance term is given by

$$\begin{aligned}
\Psi_{ij} &= E\left(f_1\left(y,\theta_i,\theta_j,\sigma\right)\middle|\boldsymbol{Y}_n\right) \\
&= \sum_z \int_{\Theta_{M_n}} \int_\sigma f_1(y,\theta_i,\theta_j,\sigma) L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}),
\end{aligned} \tag{70}$$

We begin by bounding the following term

$$\begin{aligned}
\beta_1 &= \frac{1}{D}\sum_z \int_{\sigma_n}^\infty \int_{\Theta_{M_n}} f_1(y,\theta_i,\theta_j,\sigma) L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}) \\
&\leq (H_2^*)^2 P\left(\sigma > \sigma_n \mid \boldsymbol{Y}_n\right) \\
&\leq (H_2^*)^2 \epsilon^*_{M_n},
\end{aligned} \tag{71}$$

using Lemma 7.2 of MB, where
$H_2^* = \sup_{\{y,\theta_i,\sigma\}} \left| \frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) - E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) \middle| \boldsymbol{Y}_n\right)\right| \leq \frac{2}{k}$. In fact, abusing notation a bit, we re-define $H_2^* = \frac{2}{k}$.

When $\sigma \leq \sigma_n$, for simplifying calculations, we split the range of z as $\{R_1^* \cap R_2^*\} \cup \{R_1^* \cap (R_2^*)^c\} \cup \{(R_1^*)^c \cap R_2^*\} \cup \{(R_1^*)^c \cap (R_2^*)^c\}$, where $R_2^* = \{z$: no $z_k = j\}$. The respective cardinalities are $(M_n-2)^n$, $(M_n-1)^n - (M_n-2)^n$, $(M_n-1)^n - (M_n-2)^n$ and $M_n^n - 2(M_n-1)^n + (M_n-2)^n$.

Now consider the sum $\sum_{R_1^*\cap R_2^*}\int_0^{\sigma_n}\int_{\Theta_{M_n}} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})$. In this sum none of θ_i and θ_j has been represented. Splitting the range of integration into $I_1 \cup I_2$, we obtain the following bounds:

$$\begin{aligned}
\beta_2 &= \frac{1}{D}\sum_{R_1^*\cap R_2^*}\int_{I_1} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n}) \\
&\leq (H_2^*)^2 P\left(Z \in R_1^* \cap R_2^*, \Theta_{M_n} \in E^c, \sigma \leq \sigma_n | \boldsymbol{Y}_n\right) \\
&\leq (H_2^*)^2 (M_n - 2) B_{M_n} \qquad (72)
\end{aligned}$$

and

$$\begin{aligned}
\beta_3 &= \frac{1}{D}\sum_{R_1^*\cap (R_2^*)^c}\int_{I_2} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n}) \\
&\leq (H_2^*)^2 P\left(Z \in R_1^* \cap R_2^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n | \boldsymbol{Y}_n\right) \\
&\leq (H_2^*)^2\left(1 - \frac{2}{M_n}\right)^n\left(\frac{\alpha + M_n}{\alpha}\right)^{M_n}. \qquad (73)
\end{aligned}$$

We follow the same procedure to obtain bounds for the ranges $R_1^* \cap (R_2^*)^c$ and $(R_1^*)^c \cap R_2^*$.

$$\begin{aligned}
\beta_4 &= \frac{1}{D}\sum_{z\in R_1^*\cap (R_2^*)^c}\int_{I_1} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n}) \\
&\leq (H_2^*)^2 (M_n - 1) B_{M_n}. \qquad (74)
\end{aligned}$$

and

$$\begin{aligned}
\beta_5 &= \frac{1}{D}\sum_{z\in (R_1^*)^c\cap R_2^*}\int_{I_1} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n}) \\
&\leq (H_2^*)^2 (M_n - 1) B_{M_n}. \qquad (75)
\end{aligned}$$

$$\begin{aligned}
\beta_6 &= \frac{1}{D}\sum_{z\in R_1^*\cap(R_2^*)^c}\int_{I_2} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})\\
&\le (H_2^*)^2\left(\left(1-\frac{1}{M_n}\right)^n-\left(1-\frac{2}{M_n}\right)^n\right)\left(\frac{\alpha+M_n}{\alpha}\right)^{M_n}. \qquad (76)
\end{aligned}$$

and

$$\begin{aligned}
\beta_7 &= \frac{1}{D}\sum_{z\in (R_1^*)^c\cap R_2^*}\int_{I_2} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})\\
&\le (H_2^*)^2\left(\left(1-\frac{1}{M_n}\right)^n-\left(1-\frac{2}{M_n}\right)^n\right)\left(\frac{\alpha+M_n}{\alpha}\right)^{M_n}. \qquad (77)
\end{aligned}$$

Inequalities (72)-(77) associated with $\beta_2-\beta_7$ yield

$$\begin{aligned}
&\frac{1}{D}\sum_{z\in R_1^*\cap R_2^*}\int_0^{\sigma_n}\int_{\Theta_{M_n}} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})\\
&+\frac{1}{D}\sum_{z\in R_1^*\cap (R_2^*)^c}\int_0^{\sigma_n}\int_{\Theta_{M_n}} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})\\
&+\frac{1}{D}\sum_{z\in (R_1^*)^c\cap R_2^*}\int_0^{\sigma_n}\int_{\Theta_{M_n}} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})\\
=\;& O\left((M_n-1)B_{M_n}+\left(1-\frac{2}{M_n}\right)^n\left(\frac{\alpha+M_n}{\alpha}\right)^{M_n}\right)\\
&+O\left(\left(\left(1-\frac{1}{M_n}\right)^n-\left(1-\frac{2}{M_n}\right)^n\right)\left(\frac{\alpha+M_n}{\alpha}\right)^{M_n}\right)\\
=\;& O\left(MB_{M_n}+\left(1-\frac{1}{M_n}\right)^n\left(\frac{\alpha+M_n}{\alpha}\right)^{M_n}\right).
\end{aligned}$$

(78)

Now we concentrate on the part $z\in(R_1^*)^c\cap(R_2^*)^c$. We split the range of integration of Θ_{M_n} and σ as the union of I_1^*={Either θ_i or θ_j are in $[-a-c,a+c]^c$, $\sigma\in(0,\sigma_n)$} and I_2^*={Both θ_i and θ_j are in $[-a-c,a+c]$, $\sigma\in(0,\sigma_n)$}. We then have

$$\begin{aligned}
\beta_8 &= \frac{1}{D}\sum_{z\in (R_1^*)^c\cap(R_2^*)^c}\int_{I_1^*} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})\\
&\le (H_1^*)^2B_{M_n}. \qquad (79)
\end{aligned}$$

$$
\begin{aligned}
\beta_9 &= \frac{1}{D}\sum_{z\in(R_1^*)^c\cap(R_2^*)^c}\int_{I_2^*} f_1(y,\theta_i,\theta_j,\sigma)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})\\
&= \left[\frac{1}{(\sigma_{in}(y)+k)}\phi\left(\frac{y-\theta_{in}(y)}{\sigma_{in}(y)+k}\right)-\right.\\
&\qquad \left. E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right)\right]\times\frac{1}{D}\times\\
&\quad \sum_{z\in(R_1^*)^c\cap(R_2^*)^c}\int_{I_2^*}\left[\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)-\right.\\
&\qquad \left. E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right)\right]\times L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n}),
\end{aligned}
\tag{80}
$$

applying $GMVT$, where, for each y, $\theta_{in}(y)\in(-a-c,a+c)$, and $\sigma_{in}(y)\in(0,\sigma_n)$.

Now applying $GMVT$ to the second factor of (80) we get

$$
\begin{aligned}
&\frac{1}{D}\sum_{z\in(R_1^*)^c\cap(R_2^*)^c}\int_{I_2^*}\left[\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)-\right.\\
&\qquad\qquad \left. E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right)\right] L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})\\
= &\left[\frac{1}{(\sigma_{jn}(y)+k)}\phi\left(\frac{y-\theta_{jn}(y)}{\sigma_{jn}(y)+k}\right)-E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right)\right]\\
\times &\ P\left(Z\in(R_1^*)^c\cap(R_2^*)^c,\theta_i\in(-a-c,a+c),\right.\\
&\qquad\qquad \left.\theta_j\in[-a-c,a+c],\sigma\le\sigma_n|\boldsymbol{Y}_n\right),
\end{aligned}
\tag{81}
$$

where the symbols have the same meanings as in (80).

Letting β_i^* stand for the posterior probabilities associated with β_i; $i=1,\ldots,9$, note that

$$
\begin{aligned}
\beta_9^* &= P\left(Z\in(R_1^*)^c\cap(R_2^*)^c,\theta_i\in[-a-c,a+c],\right.\\
&\qquad\qquad \left.\theta_j\in[-a-c,a+c],\sigma\le\sigma_n|\boldsymbol{Y}_n\right)\\
&= 1-\sum_{i=1}^{8}\beta_i^*,
\end{aligned}
$$

where, for each $i = 1, \ldots, 8$, $O(\beta_i^*) = O(\beta_i)$. Following the same method for obtaining (69) it can be shown that

$$\beta_9 = O\left(MB_{M_n} + \left(1 - \frac{1}{M_n}\right)^n \left(\frac{\alpha + M_n}{\alpha}\right)^{M_n}\right). \tag{82}$$

Thus, for any $(i, j) = 1, \ldots, M_n; i \neq j$, the order of the covariance term can be summarized from (71), (78), (79) and (82) as,

$$\Psi^* = \Psi_{ij} = O\left(MB_{M_n} + \left(1 - \frac{1}{M_n}\right)^n \left(\frac{\alpha + M_n}{\alpha}\right)^{M_n} + \epsilon^*_{M_n}\right). \tag{83}$$

The order of $\sum_{i=1}^{M_n} \sum_{j=1, j \neq i} Cov\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right), \frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right) \middle| \boldsymbol{Y}_n\right) = \frac{M_n(M_n-1)}{M_n^2}\Psi^*$. Taking M_n to be sufficiently large it can be ensured that $\frac{M_n(M_n-1)}{M_n^2} \cong 1$ and hence the lemma follows.

S-6. Proofs of results associated with Section 12 of MB.

S-6.1. *Proof of Lemma 12.1.*

PROOF. The proof will follow in the same way as in Lemma 6.2. The form of the likelihood now is

$$L\left(\Theta_n, z, \boldsymbol{Y}\right) = \frac{1}{\sigma^{np}} e^{-\sum_{i=1}^n \sum_{j=1}^p \frac{(Y_{ij}-\theta_{ij})^2}{2\sigma^2}} \tag{84}$$

Note that $|Y_{ij}| < a$ and $|\theta_{ij}| < c_1$ implies that $\sum_{i=1}^n \sum_{j=1}^p \frac{(Y_{ij}-\theta_{ij})^2}{2\sigma^2} = np(a + c_1)^2$. Also note that

$$\begin{aligned} & \int_{\Theta_{-i}} \int_{\theta_i \in [-a-c, a+c]^p} dH(\Theta_n) \\ \geq & \left(\frac{\alpha}{\alpha + n}\right)^n \int_{\Theta_{-i}} \int_{\theta_i \in [-a-c, a+c]^p} \prod_{l=1}^n dG_0(\theta_l) \\ = & \left(\frac{\alpha}{\alpha + n}\right)^n H_0^{np}. \end{aligned} \tag{85}$$

These observations with the same calculations as associated with Lemma 6.2 yields the required result.

□

S-6.2. *Splitting the integral* $\int_{\Theta_n}\int_0^{\sigma_n} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)$.

$$
\begin{aligned}
& \int_{\Theta_n}\int_0^{\sigma_n} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma) \\
= & \int_0^{\sigma_n}\int\int_{\theta_{i1}\in[-a-c,a+c]^c} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma) \\
& +\int_0^{\sigma_n}\int\int_{\theta_{i1}\in[-a-c,a+c]} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma) \\
= & \int_0^{\sigma_n}\int\int_{\theta_{i1}\in[-a-c,a+c]^c} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma) \\
& +\sum_{j=2}^{p}\int_0^{\sigma_n}\int\int_{W_{ij}} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma) \\
& +\int_0^{\sigma_n}\int\int_{\theta_{i1}\in[-a-c,a+c]}\cdots\int_{\theta_{ip}\in[-a-c,a+c]} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma) \\
= & \int_0^{\sigma_n}\int_{\theta_i\in E^c} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma) \\
& +\int_0^{\sigma_n}\int_{\theta_i\in E} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma),
\end{aligned}
\tag{86}
$$

where W_{ij}=$\{\theta_{i1}\in[-a-c,a+c],\ldots,\theta_{i(j-1)}\in[-a-c,a+c],\theta_{ij}\in[-a-c,a+c]^c$, rest θ_i's are in $\Re^p\}$, $j=2,\ldots,n$; E=$\{\theta_{il}\in[-a-c,a+c],\forall l$, rest θ_l's are in $\Re^p\}$; $\Re$ representing the real line.

S-6.3. *Proof of Lemma 12.2.*

PROOF. The proof follows in the same way as that of Lemma 6.3.
Note that

$$
\begin{aligned}
& P\left(\Theta_n\in E^c,\sigma\leq\sigma_n\mid\boldsymbol{Y}_n\right) \\
= & \frac{\int_\sigma\int\int_{\theta_{i1}\in[-a-c,a+c]^c} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)}{\int_\sigma\int_{\Theta_n} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)} \\
& +\frac{\sum_{j=2}^{p}\int_\sigma\int\int_{W_{ij}} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)}{\int_\sigma\int_{\Theta_n} L(\Theta_n,\boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma)}
\end{aligned}
\tag{87}
$$

In the same way as in Lemma 6.3 it can be shown that each of the p component probabilities is of the order $B_n=\frac{\alpha+n}{\alpha}e^{-\frac{c^2}{4\sigma_n^2}}$. Hence, it is easy to see that

$$P\left(\Theta_n\in E^c,\sigma\leq\sigma_n\mid\boldsymbol{Y}_n\right)=O\left(pB_n\right).$$

□

S-7. Splitting of $\int_0^{\sigma_n}\int_{\Theta_{M_n}} L\left(\Theta_{M_n}, z, \boldsymbol{Y}_n\right) dH(\Theta_n)dG_n(\sigma)$ in the case of the "large p small n problem" associated with the SB model. As before, for $z \in R_1^*$, denote by d the number of θ_l's present in likelihood, Θ_z, the set of θ_l's that are present in the likelihood, and $\theta_1, \ldots, \theta_d$, the θ_l's present in the likelihood. Here also for $z \in R_1^*$, we split the integral $\int_0^{\sigma_n}\int_{\Theta_{M_n}} L\left(\Theta_{M_n}, z, \boldsymbol{Y}_n\right) dH(\Theta_n)dG_n(\sigma)$ as

$$
\begin{aligned}
& \int_0^{\sigma_n}\int_{\Theta_{M_n}} L\left(\Theta_{M_n}, z, \boldsymbol{Y}_n\right) dH(\Theta_n)dG_n(\sigma) \\
= & \int_0^{\sigma_n}\int\int_{\theta_{11}\in[-a-c,a+c]} L\left(\Theta_{M_n}, z, \boldsymbol{Y}_n\right) dH(\Theta_{M_n})dG_n(\sigma) \\
& +\sum_{j=1}^{d}\sum_{l=1}^{p}\int_0^{\sigma_n}\int\int_{W_{jl}} L\left(\Theta_{M_n}, z, \boldsymbol{Y}_n\right) dH(\Theta_{M_n})dG_n(\sigma) \\
& +\int_0^{\sigma_n}\int_{\Theta_{M_n}\in E} L\left(\Theta_{M_n}, z, \boldsymbol{Y}_n\right) dH(\Theta_{M_n})dG_n(\sigma) \\
= & \int_0^{\sigma_n}\int_{\Theta_{M_n}\in E^c} L\left(\Theta_{M_n}, z, \boldsymbol{Y}_n\right) dH(\Theta_{M_n})dG_n(\sigma) \\
& +\int_0^{\sigma_n}\int_{\Theta_{M_n}\in E} L\left(\Theta_{M_n}, z, \boldsymbol{Y}_n\right) dH(\Theta_{M_n})dG_n(\sigma),
\end{aligned}
\tag{88}
$$

where $E=\{$all θ_l's in likelihood are in $[-a-c, a+c]$, rest are in $\Re^p\}$, $W_{jl}=\{\theta_{j1}\in[-a-c,a+c], \ldots, \theta_{j(l-1)}\in[-a-c,a+c], \theta_{jl}\in[-a-c,a+c]^c$, rest θ_{st}'s are in $\Re^p\}$.

S-8. Proofs of results associated with Section 13 of MB.

S-8.1. *Proof of Theorem 13.1.*

PROOF. Let $\boldsymbol{Y}_n \in \boldsymbol{S}_n$, as before. Now note that

$$
\begin{aligned}
& E\left(\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\bigg|\boldsymbol{Y}_n\right) \\
= & \frac{1}{D}\int_\sigma\int_{\Theta_n}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_n, \boldsymbol{Y}_n)dH(\Theta_n)dG_n(\sigma) \\
\leq & H_1,
\end{aligned}
\tag{89}
$$

where $H_1 = \sup_{\{y,\theta_i,\sigma\}} \left\{ \frac{1}{(\sigma+k)} \phi\left(\frac{y-\theta_i}{\sigma+k}\right) \right\} = \frac{1}{k}$. As a result,

$$\begin{aligned} & \frac{1}{\alpha+n} \sum_{i=1}^{n} E\left(\frac{1}{(\sigma+k)} \phi\left(\frac{y-\theta_i}{\sigma+k} \right) \middle| \boldsymbol{Y}_n \right) \\ \leq & \frac{n}{\alpha+n} H_1 \\ \rightarrow & 0. \end{aligned} \tag{90}$$

Now consider

$$\begin{aligned} & A_n \\ = & \frac{1}{D} \int_{\theta_{n+1}} \int_{\Theta_n} \int_{\sigma} \frac{1}{(\sigma+k)} e^{-\frac{(y-\theta_{n+1})^2}{2(\sigma+k)^2}} L(\Theta_n, \boldsymbol{Y}_n) dG_0(\theta_{n+1}) dH(\Theta_n) dG_n(\sigma) \\ = & \frac{1}{D} \int_{\theta_{n+1}} \int_{\Theta_n} \int_{\sigma<\sigma_n} \frac{1}{(\sigma+k)} e^{-\frac{(y-\theta_{n+1})^2}{2(\sigma+k)^2}} L(\Theta_n, \boldsymbol{Y}_n) dG_0(\theta_{n+1}) dH(\Theta_n) dG_n(\sigma) \\ = & \frac{1}{D} \int_{\theta_{n+1}} \int_{\Theta_n} \int_{\sigma>\sigma_n} \frac{1}{(\sigma+k)} e^{-\frac{(y-\theta_{n+1})^2}{2(\sigma+k)^2}} L(\Theta_n, \boldsymbol{Y}_n) dG_0(\theta_{n+1}) dH(\Theta_n) dG_n(\sigma) \\ = & W_1 + W_2 \ \text{(say).} \end{aligned} \tag{91}$$

$$\begin{aligned} W_2 = & \frac{1}{D} \int_{\theta_{n+1}} \int_{\Theta_n} \int_{\sigma>\sigma_n} \frac{1}{(\sigma+k)} e^{-\frac{(y-\theta_{n+1})^2}{2(\sigma+k)^2}} L(\Theta_n, \boldsymbol{Y}_n) dG_0(\theta_{n+1}) dH(\Theta_n) dG_n(\sigma) \\ \leq & H_1 P\left(\sigma > \sigma_n | \boldsymbol{Y}_n\right). \end{aligned}$$

Thus, by Lemma 6.2 of MB,

$$W_2 = O(\epsilon_n^*). \tag{92}$$

As regards W_1, an application of $GMVT$ yields,

$$\begin{aligned} W_1 = & \frac{1}{D} \int_{\theta_{n+1}} \int_{\Theta_n} \int_{\sigma<\sigma_n} \frac{1}{(\sigma+k)} e^{-\frac{(y-\theta_{n+1})^2}{2(\sigma+k)^2}} L(\Theta_n, \boldsymbol{Y}_n) dG_0(\theta_{n+1}) dH(\Theta_n) dG_n(\sigma) \\ = & \int_{\theta_{n+1}} \frac{1}{(\sigma_n^*(y)+k)} e^{-\frac{(y-\theta_{n+1})^2}{2(\sigma_n^*(y)+k)^2}} dG_0(\theta_{n+1}) \times P\left(\sigma < \sigma_n | \boldsymbol{Y}_n\right), \end{aligned}$$

DCT ensures that

$$\begin{aligned} & \int_{\theta_{n+1}} \frac{1}{(\sigma_n^*(y)+k)} e^{-\frac{(y-\theta_{n+1})^2}{2(\sigma_n^*(y)+k)^2}} dG_0(\theta_{n+1}) \\ & \rightarrow \int_{\theta_{n+1}} \frac{1}{k} e^{-\frac{(y-\theta_{n+1})^2}{2k^2}} dG_0(\theta_{n+1}). \end{aligned} \tag{93}$$

It then follows from (93) and the fact that $P\left(\sigma < \sigma_n | \boldsymbol{Y}_n\right) \rightarrow 1$, that

$$W_1 \rightarrow \int_{\theta_{n+1}} \frac{1}{k} e^{-\frac{(y-\theta_{n+1})^2}{2k^2}} dG_0(\theta_{n+1}). \tag{94}$$

Finally, (92) and (94) guarantee Theorem 13.1. □

S-8.2. *Proof of Lemma 13.2.*

PROOF.

$$\begin{aligned} P(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n | \boldsymbol{Y}_n) &= P(Z \in R_1^*, \Theta_{M_n} \in E | \boldsymbol{Y}_n) \\ &\quad - P(Z \in R_1^*, \Theta_{M_n} \in E, \sigma > \sigma_n | \boldsymbol{Y}_n) \end{aligned} \tag{95}$$

We first obtain a lower bound for $P(Z \in R_1^*, \Theta_{M_n} \in E | \boldsymbol{Y}_n)$.

$$\begin{aligned} & P\left(Z \in R_1^*, \Theta_{M_n} \in E | \boldsymbol{Y}_n\right) \\ = & \frac{\sum_{z \in R_1^*} \int_{\sigma \leq \sigma_n} \int_{\Theta_z \in E} \int L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})}{\sum_z \int_\sigma \int_{\Theta_{M_n}} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})} \\ = & \frac{\sum_{z \in R_1^*} N}{\sum_z D}, \end{aligned} \tag{96}$$

where $N = \int_{\sigma \leq \sigma_n} \int_{\Theta_z \in E} \int L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})$ and
$D = \int_\sigma \int_{\Theta_{M_n}} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})$.

Denote $E^* = \{\theta_j \in [-a-c, a+c] \cap [\bar{y}_j - k_n, \bar{y}_j + k_n], j = 1, \ldots, d; \theta_j \in (-\infty, \infty), j = d+1, \ldots, M_n\}$ and $E_j = \{\theta_j \in [-a-c, a+c] \cap [\bar{y}_j - k_n, \bar{y}_j + k_n]\}$ for $j = 1, \ldots, d$. Let k_n be a sequence of constants such that $k_n \rightarrow \infty$.
Note that,

$$\begin{aligned} N \geq & \int_{\Theta_{M_n} \in E^*} \int_{nk_n}^{2nk_n} \frac{1}{\sigma^n} e^{-\frac{\sum_{j=1}^d \sum_{t: z_t = j} (y_t - \bar{y}_j)^2}{2\sigma^2}} \\ & \times e^{-\frac{\sum_{j=1}^d n_j (\bar{y}_j - \theta_j)^2}{2\sigma^2}} dH(\Theta_{M_n}) dG_n(\sigma) \\ \geq & \left(\frac{1}{2nk_n}\right)^n e^{-\frac{C_n^{(2)}}{8n^2 k_n^2}} \times e^{-\frac{1}{2n}} \times \left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} \times \prod_{j=1}^d G_0\left(E_j\right) \times O(\epsilon_n) \end{aligned} \tag{97}$$

assuming $\int_{nk_n}^{2nk_n} dG_n(\sigma) = O(\epsilon_n)$.

Since $k_n \to \infty$, as $n \to \infty$, $G_0(E_j) \sim G_0((-\infty,\infty)) = 1$ and

$$N \gtrsim \left(\frac{1}{2nk_n}\right)^n e^{-\frac{C_n^{(2)}}{8n^2k_n^2}} \times e^{-\frac{1}{2n}} \times \left(\frac{\alpha}{\alpha+M_n}\right)^{M_n} \times O(\epsilon_n). \tag{98}$$

Now,

$$e^{-\frac{\sum_{j=1}^{M_n}\sum_{t:z_t=j}(y_t-\bar{y}_j)^2}{2\sigma^2}} \times \frac{1}{\sigma^n} e^{-\frac{\sum_{j=1}^{M_n} n_j(\bar{y}_j-\theta_j)^2}{2\sigma^2}} \leq 1 \times \left(\frac{n}{C_n^{(1)}}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}, \tag{99}$$

for $0 < \sigma < \infty$. This implies

$$D \leq \left(\frac{n}{C_n^{(1)}}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}. \tag{100}$$

Since $C_n^{(1)} \sim C_n^{(2)} \sim C_n$ for large n, let us obtain the condition under which

$$\left(\frac{1}{2nk_n}\right)^n \times e^{-\frac{C_n}{8n^2k_n^2}} \times O(\epsilon_n) \geq \left(\frac{n}{C_n}\right)^{\frac{n}{2}} e^{-\frac{n}{2}} \tag{101}$$

Let $k_n = r_nC_n$, where $r_n \to 0$ and $r_nC_n \to \infty$.
Then,

$$\begin{aligned}
&\left(\frac{1}{2nk_n}\right)^n \times e^{-\frac{C_n}{8n^2k_n^2}} \times O(\epsilon_n) \geq \left(\frac{n}{C_n}\right)^{\frac{n}{2}} e^{-\frac{n}{2}} \\
\Leftrightarrow\quad &\left(\frac{1}{2nr_nC_n}\right)^n \times e^{-\frac{C_n}{8n^2r_n^2C_n^2}} \times O(\epsilon_n) \geq \left(\frac{n}{C_n}\right)^{\frac{n}{2}} e^{-\frac{n}{2}} \\
\Leftrightarrow\quad &-n\log(C_n) + \frac{n}{2}\log(C_n) - \frac{C_n}{8n^2r_n^2C_n^2} \\
&\qquad \geq \frac{n}{2}\log(n) - \frac{n}{2} + n\log(2n) + n\log(r_n) - O(\log(\epsilon_n)) \\
\Leftrightarrow\quad &\frac{n}{2}\log(C_n) + \frac{1}{8n^2r_n^2C_n^2} \\
&\qquad \leq -\frac{n}{2}\log(n) + \frac{n}{2} - n\log(2) - n\log(n) - n\log(r_n) - O\left(\log\left(\frac{1}{\epsilon_n}\right)\right) \\
\Leftrightarrow\quad &\frac{n}{2}\log(C_n) + \frac{1}{8n^2r_n^2C_n^2} \\
&\qquad \leq n\left(-\frac{3}{2}\log(n) - \log(r_n) + \frac{1}{2} - \log(2)\right) - O\left(\log\left(\frac{1}{\epsilon_n}\right)\right)
\end{aligned} \tag{102}$$

In the R.H.S of the equality (102) the term $\left(\frac{1}{2} - \log(2)\right)$ is a constant and ϵ_n is chosen independently. So, we can choose r_n sufficiently small such that term $n\left(-\log(n) - \log(r_n)\right)$ dominates the other terms.

Let $C_n = \frac{1}{r_n^s n^2}$, where $s > 2$.

Then $k_n = r_n C_n = \frac{1}{r_n^{s-1} n^2} \to \infty$, for r_n going to zero at a sufficiently fast rate.

Also, $n^2 r_n^2 C_n = \frac{n^2 r_n^2}{r_n^s n^2} = \frac{1}{r_n^{s-2}} \to \infty$, for $s > 2$.

And,

$$\frac{n}{2}\log(C_n) = -\frac{ns}{2}\log(r_n) - n\log(n) < -n\log(r_n) - n\log(n). \tag{103}$$

So, for $C_n = O\left(\frac{1}{r_n^s n^2}\right)$; $s > 2$, if r_n is fixed to be sufficiently small such that for large n, $\frac{1}{8n^2 r_n^2 C_n^2} \approx 0$, and $O\left(\log\left(\frac{1}{\epsilon_n}\right)\right) \prec -n\log(r_n) - n\log(n)$. Then, as $n \to \infty$, (102) holds, and

$$\left(\frac{1}{2nk_n}\right)^n \times e^{-\frac{C_n}{8n^2 k_n^2}} \times O(\epsilon_n) \gtrsim \left(\frac{n}{C_n}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}$$

Hence, it follows that

$$\begin{aligned}
& P\left(Z \in R_1^*, \Theta_{M_n} \in E|\boldsymbol{Y}_n\right) \\
\gtrsim \quad & \frac{(M_n - 1)^n \left(\frac{1}{2nk_n}\right)^n e^{-\frac{C_n^{(2)}}{8n^2 k_n^2}} \times e^{-\frac{1}{2n}} \times \left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} \times O(\epsilon_n)}{M_n^n \left(\frac{n}{C_n^{(1)}}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}} \\
\gtrsim \quad & \left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} \left(1 - \frac{1}{M_n}\right)^n. \qquad (104)
\end{aligned}$$

Now we obtain an upper bound for $P\left(Z \in R_1^*, \Theta_{M_n} \in E, \sigma > \sigma_n|\boldsymbol{Y}_n\right)$.

$$\begin{aligned} & P\left(Z \in R_1^*, \Theta_{M_n} \in E, \sigma > \sigma_n | \boldsymbol{Y}_n\right) \\ = & \frac{\sum_{z \in R_1^*} \int_{\sigma_n}^{\infty} \int_{\Theta_z \in E} \int L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})}{\sum_z \int_{\sigma} \int_{\Theta_{M_n}} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})} \\ = & \frac{\sum_{z \in R_1^*} N}{\sum_z D} \\ \leq & \frac{\sum_{z \in R_1^*} N}{\sum_{z \in R_1^*} D}, \end{aligned} \tag{105}$$

where $N = \int_{\sigma_n}^{\infty} \int_{\Theta_z \in E} \int L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})$ and $D = \int_{\sigma} \int_{\Theta_{M_n}} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n})$.

Note that, in the same as we have obtained equation (99), it can shown that

$$N \leq \left(\frac{n}{C_n^{(1)}}\right)^{\frac{n}{2}} e^{-\frac{n}{2}} \times O(\epsilon_n). \tag{106}$$

$$\begin{aligned} D & \geq \int_{\sigma > \sigma_n} \int_{\Theta_{M_n}} L(\Theta_{M_n}, z, \boldsymbol{Y}_n) dG_n(\sigma) dH(\Theta_{M_n}) \\ & \geq \left(\frac{1}{2nk_n}\right)^n e^{-\frac{C_n^{(2)}}{8n^2k_n^2}} \times e^{-\frac{1}{2n}} \left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} \times \prod_{j=1}^{d} G_0\left(E_j\right) \times O(\epsilon_n) \\ & \gtrsim \left(\frac{1}{2nk_n}\right)^n e^{-\frac{C_n^{(2)}}{8n^2k_n^2}} \times e^{-\frac{1}{2n}} \left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} \times O(\epsilon_n), \end{aligned} \tag{107}$$

since for large n, $G_0\left(E_j\right) \sim G_0\left((-\infty, \infty)\right) = 1$.

Since $C_n^{(2)} \sim C_n^{(2)} \sim C_n$ it follows that

$$\frac{\sum_{z \in R_1^*} N}{\sum_z D} \leq \frac{\left(\frac{n}{C_n}\right)^{\frac{n}{2}} e^{-\frac{n}{2}} \times O(\epsilon_n)}{\left(\frac{1}{2nk_n}\right)^n e^{-\frac{C_n}{8n^2k_n^2}} \times e^{-\frac{1}{2n}} \left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} \times O(\epsilon_n)} \tag{108}$$

Choose C_n such that

$$\left(\frac{1}{2nk_n}\right)^n e^{-\frac{C_n}{8n^2k_n^2}} \times O(\epsilon_n) \geq \left(\frac{n}{C_n}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}, \tag{109}$$

which is exactly the same condition as in the last case of the lower bounds. So, as $n \to \infty$,

$$P\left(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \geq \sigma_n\right) \lesssim \left(\frac{\alpha + M_n}{\alpha}\right)^{M_n} \left(1 - \frac{1}{M_n}\right)^n \times O(\epsilon_n). \tag{110}$$

Hence,

$$\begin{aligned}
& P\left(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n | \boldsymbol{Y}_n\right) \\
= \; & P\left(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n | \boldsymbol{Y}_n\right) - P\left(Z \in R_1^*, \Theta_{M_n} \in E, \sigma > \sigma_n | \boldsymbol{Y}_n\right) \\
\gtrsim \; & \left[\left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} - \left(\frac{\alpha + M_n}{\alpha}\right)^{M_n} \times O(\epsilon_n)\right] \left(1 - \frac{1}{M_n}\right)^n .
\end{aligned} \tag{111}$$

We must have

$$\begin{aligned}
& \left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} \gtrsim \left(\frac{\alpha + M_n}{\alpha}\right)^{M_n} \times O(\epsilon_n) \\
\Leftrightarrow \quad & O(\epsilon_n) \lesssim \left(\frac{\alpha}{\alpha + M_n}\right)^{2M_n}
\end{aligned} \tag{112}$$

Using L' Hospital's rule it can be shown that $\left(\frac{\alpha}{\alpha+M_n}\right)^{2M_n} \to 1$, if $M_n = n^b$, $\alpha = n^{\omega}$, $\omega > 0$, $b > 0$ and $\omega - b > b$. Since $\epsilon_n \to 0$, for large n, (112) holds and does not contradict the assumptions regarding ϵ_n.

Summing up all the results we have,

$$P\left(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n | \boldsymbol{Y}_n\right) \gtrsim \left(\frac{\alpha}{\alpha + M_n}\right)^{M_n} \left(1 - \frac{1}{M_n}\right)^n , \tag{113}$$

for $C_n = O\left(\frac{1}{r_n^s n^2}\right)$, $s > 2$.

□

S-8.3. *Proof of Theorem 13.3.* Consider the integral

$$
\begin{aligned}
&\int_{\Theta_{M_n}\in E}\int_0^{\sigma_n}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)\\
=&\ \frac{\alpha}{\alpha+M_n-1}\int_{\Theta_{M_n}\in E}\int_0^{\sigma_n}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\\
&\qquad\times L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_0(\theta_i)dH_{-i}(\Theta_{-iM_n})dG_n(\sigma)\\
&+\frac{1}{\alpha+M_n-1}\sum_{j=1,j\neq i}^{M_n}\int_{\Theta_{-iM_n}\in E_{-i}}\int_0^{\sigma_n}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)\\
&\qquad\times L(\Theta_{-iM_n},z,\boldsymbol{Y}_n)dH_{-i}(\Theta_{-iM_n})dG_n(\sigma),
\end{aligned}
\tag{114}
$$

using the Polya urn representation of $H(\Theta_{M_n})$, given by

$$
H(\Theta_{M_n})=\left[\frac{\alpha}{\alpha+M_n-1}G_0(\theta_i)+\frac{1}{\alpha+M_n-1}\sum_{j=1,j\neq i}^{M_n}\delta_{\theta_j}(\theta_i)\right]\times H_{-i}(\Theta_{-iM_n}),
\tag{115}
$$

where $\Theta_{-iM_n}=\Theta_{M_n}\setminus\theta_i$ and $H_{-i}(\Theta_{-iM_n})$ is the joint distribution of Θ_{-iM_n} and E_{-i} is the set E excluding θ_i.

Let $D=\sum_z\int_\sigma\int_{\Theta_{M_n}}L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_n(\sigma)dH(\Theta_{M_n})$. Note that for $z\in R_1^*$, θ_i is not present in likelihood and hence $\{\Theta_{M_n}\in E\}=\{-\infty<\theta_i<\infty\}\cap\{\Theta_{-iM_n}\in E_{-i}\}$. Then,

$$
\begin{aligned}
&\frac{1}{D}\sum_{z\in R_1^*}\int_{\Theta_{M_n}\in E}\int_0^{\sigma_n}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)\\
&\qquad\times L(\Theta_{M_n},z,\boldsymbol{Y}_n)dG_0(\theta_i)dH_{-i}(\Theta_{-iM_n})dG_n(\sigma)\\
=&\ \frac{1}{D}\sum_{z\in R_1^*}\int_{\Theta_{-iM_n}\in E_{-i}}\int_{\theta_i}\int_0^{\sigma_n}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)dG_0(\theta_i)\\
&\qquad\times L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH_{-i}(\Theta_{-iM_n})dG_n(\sigma)\\
=&\ \int_{\theta_i}\frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y-\theta_i}{\sigma_n^*(y)+k}\right)dG_0(\theta_i)\\
&\qquad\times P_{M_n-1}\left(Z\in R_1^*,\Theta_{-iM_n}\in E_{-i},\sigma\leq\sigma_n\middle|\boldsymbol{Y}_n\right),
\end{aligned}
\tag{116}
$$

where $P_{M_n-1}(\cdot|\boldsymbol{Y}_n)$ is the posterior probability when the mixture model has M_n-1 components.

It can shown that exactly under the same conditions as Lemma 13.2, $P_{M_n-1}\left(Z\in R_1^*,\Theta_{-iM_n}\in E_{-i},\sigma\leq\sigma_n\middle|\boldsymbol{Y}_n\right)$ has the same lower bound with only M_n replaced with M_n-1. Using L' Hospital's rule it can be easily shown that $P_{M_n-1}\left(Z\in R_1^*,\Theta_{-iM_n}\in E^*,\sigma\leq\sigma_n\middle|\boldsymbol{Y}_n\right)$ also converges to 1.

DCT ensures that

$$\int_{\theta_i}\frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y-\theta_i}{\sigma_n^*(y)+k}\right)dG_0(\theta_i)\to\int_{\theta_i}\frac{1}{k}e^{-\frac{(y-\theta_i)^2}{2k^2}}dG_0(\theta_i),\tag{117}$$

where the symbols have the same meanings as in (117).

Again,

$$\begin{aligned}&\frac{1}{\alpha+M_n-1}\frac{1}{D}\sum_{z\in R_1^*}\sum_{j=1,j\neq i}^{M_n}\int_{\Theta_{-iM_n}\in E_{-i}}\int_0^{\sigma_n}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)\\&\qquad\qquad\times L(\Theta_{-iM_n},z,\boldsymbol{Y}_n)dH_{-i}(\Theta_{-iM_n})dG_n(\sigma)\\\leq\quad& H_1\times\frac{1}{\alpha+M_n-1}\\&\times\sum_{j=1,j\neq i}^{M_n}\frac{1}{D}\sum_{z\in R_1^*}\int_{\Theta_{-iM_n}\in E_{-i}}\int_0^{\sigma_n}L(\Theta_{-iM_n},z,\boldsymbol{Y}_n)dH_{-i}(\Theta_{-iM_n})dG_n(\sigma)\\\leq\quad& H_1\times\frac{M_n-1}{\alpha+M_n-1},\end{aligned}\tag{118}$$

where $H_1=\sup_{y,\theta_i,\sigma}\left\{\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_j}{\sigma+k}\right)\right\}=\frac{1}{k}$. Note that for $\alpha\succ O(M_n)$, $\frac{\alpha}{\alpha+M_n-1}\to 1$ and $\frac{M_n-1}{\alpha+M_n-1}\to 0$.

From (117) and (118) we conclude that for $\boldsymbol{Y}_n\in\boldsymbol{S}_n$,

$$\begin{aligned}&\int_{\Theta_{M_n}\in E}\int_0^{\sigma_n}\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma)\\&\to\int_{\theta_i}\frac{1}{k}e^{-\frac{(y-\theta_i)^2}{2k^2}}dG_0(\theta_i),\end{aligned}\tag{119}$$

as $n\to\infty$.

S-9. Overview of asymptotic calculations associated with Section 14 of MB. It is easy to see that the upper bounds of the probabilites given in Lemmas 7.2–7.5 remain the same for this modified model. For the modified SB model the likelihood function $L(\Theta_{M_n}, z, \boldsymbol{Y}_n, \Pi)$ is given by

$$L(\Theta_{M_n}, z, \boldsymbol{Y}_n, \Pi) = \prod_{\ell=1}^{M_n} \pi_\ell^{n_\ell+\beta_\ell-1} \prod_{j=1}^{M_n} \frac{1}{\sigma^{n_j}} e^{-\frac{1}{2}\sum_{t:z_t=j}\left(\frac{Y_t-\theta_j}{\sigma}\right)^2}. \tag{120}$$

From the form (120) it is clear that given z, the posterior of Π is independent of Θ_{M_n}. Hence, it is easy to see that, same calculations as in Lemma 7.2 yield the following bounds for the modified model:

$$N \leq \frac{1}{(\sigma_n)^n} P(\sigma > \sigma_n) \sum_z \int_\Pi \prod_{\ell=1}^{M_n} \pi_\ell^{n_\ell+\beta_\ell-1} d\Pi$$

and

$$D \geq \frac{\exp\left(\frac{-n(a+c_1)^2}{2(b_n)^2}\right)}{(b_n)^n} P(b_n < \sigma \leq \sigma_n) \left(\frac{\alpha}{\alpha+M_n}\right)^{M_n} H_0^{M_n} \sum_z \int_\Pi \prod_{\ell=1}^{M_n} \pi_\ell^{n_\ell+\beta_\ell-1} d\Pi.$$

Hence, the upper bound for $P\left(\sigma > \sigma_n | \boldsymbol{Y}_n\right)$ does not change. Similarly, the same argument shows that the upper bounds remain the same for all the probabilities except for $P(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n | \boldsymbol{Y}_n)$. For $P(Z \in R_1^*, \Theta_{M_n} \in E, \sigma \leq \sigma_n | \boldsymbol{Y}_n)$, the same calculations as in Lemma 3.3 show that

$$\begin{aligned} N \quad &\leq \quad \left(\frac{1}{\sigma_n}\right)^n e^{-\frac{C_n^{(1)}}{2\sigma_n^2}} \times G_0([-a-c, a+c]) \times O(1-\epsilon_n) \\ &\qquad \times \sum_{z \in R_1^*} \int_\Pi \prod_{\ell=1}^{M_n} \pi_\ell^{n_\ell+\beta_\ell-1} d\Pi. \end{aligned} \tag{121}$$

Similarly,

$$\begin{aligned} D \quad &\geq \quad \left(\frac{1}{2nk_n}\right)^n \times e^{-\frac{C_n^{(2)}}{8n^2k_n^2}} \times e^{-\frac{1}{2n}} \times \left(\frac{\alpha}{\alpha+M_n}\right)^{M_n} \\ &\qquad \times \prod_{j=1}^d G_0([\bar{Y}_j - k_n, \bar{Y}_j + k_n]) \times O(\epsilon_n) \\ &\qquad \times \sum_z \int_\Pi \prod_{\ell=1}^{M_n} \pi_\ell^{n_\ell+\beta_\ell-1} d\Pi. \end{aligned} \tag{122}$$

Since

$$\frac{\sum_{z\in R_1^*}\int_\Pi \prod_{\ell=1}^{M_n} \pi_\ell^{n_\ell+\beta_\ell-1} d\Pi}{\sum_z \int_\Pi \prod_{\ell=1}^{M_n} \pi_\ell^{n_\ell+\beta_\ell-1} d\Pi} = P(Z\in R_1^*) = \frac{(M_n-1)^n}{M_n^n},$$

the upper bound remains the same as before.

It can also be shown that $E\left(\hat{f}_{SB}^*(y\mid\Theta_{M_n},\Pi,\sigma)\right)$ converges to $\frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)$. We will split the expectation in the same way as in Theorem 3.5 into S_1,S_2,S_3,S_4,S_5, with the integrand $\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)$ replaced with $\sum_{i=1}^{M_n}\pi_i\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right)$. The upper bounds of S_i for $i\neq 4$, will be same. We illustrate this with S_1; for the others the same arguments will hold.

$$\begin{aligned} S_1 &= \frac{1}{D}\sum_{R_1^*}\int_\Pi\int_{I_1}\sum_{i=1}^{M_n}\pi_i\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) \\ &\qquad\times L(\Theta_{M_n},z,\boldsymbol{Y}_n,\Pi)dH(\Theta_{M_n})dG_n(\sigma)d\Pi \\ &\leq H_1^*\frac{1}{D}\sum_{R_1^*}\int_\Pi\int_{I_1}\sum_{i=1}^{M_n}\pi_i L(\Theta_{M_n},z,\boldsymbol{Y}_n,\Pi)dH(\Theta_{M_n})dG_n(\sigma)d\Pi \\ &= H_1^*\frac{1}{D}\sum_{R_1^*}\int_{I_1} L(\Theta_{M_n},z,\boldsymbol{Y}_n)dH(\Theta_{M_n})dG_n(\sigma), \end{aligned}$$

using the fact that $\sum_{i=1}^{M_n}\pi_i=1$. Hence S_1 has same order as $P(Z\in R_1^*,\Theta_{M_n}\in E,\sigma\leq\sigma_n|\boldsymbol{Y}_n)$ for the modified model also.

To investigate the form of the density where the modified SB model converges to, note that

$$\begin{aligned} S_4 &= \frac{1}{D}\sum_{(R_1^*)^c}\int_\Pi\int_{I_4}\sum_{i=1}^{M_n}\pi_i\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) \\ &\qquad\times L(\Theta_{M_n},z,\boldsymbol{Y}_n,\Pi)dH(\Theta_{M_n})dG_n(\sigma)d\Pi \\ &= \frac{1}{D}\int_\Pi\int_{I_4}\sum_{i=1}^{M_n}\pi_i\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) \\ &\qquad\times\sum_{(R_1^*)^c} L(\Theta_{M_n},z,\boldsymbol{Y}_n,\Pi)dH(\Theta_{M_n})dG_n(\sigma)d\Pi. \end{aligned}$$

(123)

For each i, using $GMVT$ we get

$$\begin{aligned}
& \frac{1}{D}\int_{\Pi}\int_{I_4}\pi_i\frac{1}{(\sigma+k)}\phi\left(\frac{y-\theta_i}{\sigma+k}\right) \\
& \qquad \times\sum_{(R_1^*)^c} L(\Theta_{M_n},z,\boldsymbol{Y}_n,\Pi)dH(\Theta_{M_n})dG_n(\sigma)d\Pi \\
= \quad & \frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y-\theta_n^*(y)}{\sigma_n^*(y)+k}\right) \\
& \qquad \times\frac{1}{D}\int_{\Pi}\int_{I_4}\pi_i\sum_{(R_1^*)^c} L(\Theta_{M_n},z,\boldsymbol{Y}_n,\Pi)dH(\Theta_{M_n})dG_n(\sigma)d\Pi,
\end{aligned} \tag{124}$$

where, for every y, $\theta_n^*(y)\in(-a-c,a+c)$, and $\sigma_n^*(y)\in(0,\sigma_n)$.

Hence, S_4 given by (123) becomes

$$\begin{aligned}
S_4 \quad = \quad & \frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y-\theta_n^*(y)}{\sigma_n^*(y)+k}\right) \\
& \qquad \times\frac{1}{D}\int_{\Pi}\int_{I_4}\sum_{(R_1^*)^c} L(\Theta_{M_n},z,\boldsymbol{Y}_n,\Pi)dH(\Theta_{M_n})dG_n(\sigma)d\Pi, \\
= \quad & \frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y-\theta_n^*(y)}{\sigma_n^*(y)+k}\right) \\
& \qquad \times P\left((R_1^*)^c,I_4|\boldsymbol{Y}_n\right),
\end{aligned} \tag{125}$$

again using the fact that $\sum_{i=1}^{M_n}\pi_i=1$.

Since it is already shown in Section 7 of MB that $\frac{1}{(\sigma_n^*(y)+k)}\phi\left(\frac{y-\theta_n^*(y)}{\sigma_n^*(y)+k}\right)\rightarrow \frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)$ and $P\left((R_1^*)^c,I_4|\boldsymbol{Y}_n\right)\rightarrow 1$, it follows that $S_4\rightarrow\frac{1}{k}\phi\left(\frac{y-\theta^*(y)}{k}\right)$. With very minor adjustments to the proof of Theorem 7.6, here it can be proved that the EW model and the modified SB model converge to the same distribution.

Hence, it is easy to see that the order of the $MISE$ of the modified SB model will remains the same as the previous version of the SB model.

S-10. Prior predictive convergence rates.

S-10.1. *Models under consideration.*

S-10.1.1. *Assumptions on the true distribution.* Let f_0 denote the true distribution. We assume that f_0 is continuously differentiable of all orders. Thus, the assumptions on f_0 for the prior calculations are different from those associated with the posterior calculations.

S-10.1.2. *EW model.* We denote $\Theta_n = \{\theta_1, \ldots, \theta_n\}$ and assume $\theta_i \overset{iid}{\sim} F$ and $F \sim D(\alpha G_0)$, where $D(\alpha G_0)$ denotes the Dirichlet process with central measure G_0 and hyperparameter α. Then, according to the EW model, the prior predictive density at the point y has the following form:

$$\hat{f}_{EW}(y \mid \Theta_n, h) = \frac{\alpha}{\alpha+n}A + \frac{1}{\alpha+n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{y-\theta_i}{h}\right), \tag{126}$$

where we assume the kernel $K(\cdot)$ to be symmetric around zero, $A = \int_\theta \frac{1}{h}K\left(\frac{y-\theta}{h}\right)G_0(\theta)d\theta$. In our work on calculating prior rates, we determine the optimum value of h by minimizing the rate of convergence.

S-10.1.3. *SB model.* The SB model assumes that the density at the point y is given by,

$$\hat{f}_{SB}(y \mid \Theta_{M_n}, h) = \frac{1}{M_n}\sum_{l=1}^{M_n}\frac{1}{h}K\left(\frac{y-\theta_l}{h}\right), \tag{127}$$

where for each l, $\theta_l \sim F$; $F \sim D(\alpha G_0)$, $\Theta_{M_n} = \{\theta_1, \ldots, \theta_{M_n}\}$; M_n being the maximum number of distinct components the mixture model can have. In both EW and SB models, we confine attention to the case where $K(\cdot)$ is chosen to be the $N(0,1)$ density.

S-10.2. *Measure of divergence.* To study the rates of convergence we need to define a measure of divergence. In our study we use the $MISE$ as the measure of divergence, defined as,

$$\begin{aligned} MISE &= \int_y E(\hat{f}(y \mid \Theta) - f(y))^2 dy \\ &= \int_y \int_\Theta \left\{\hat{f}(y \mid \Theta) - f(y)\right\}^2 dH(\Theta)dy, \end{aligned} \tag{128}$$

where the expectation is taken over Θ, the set of parameters of the model and $H(\cdot)$ is the joint distribution of Θ obtained by integrating out the unknown random measure F. Since F is a Dirichlet process in our case, $H(\cdot)$ is composed of the Polya urn distributions. As mentioned before, h will be determined by minimizing $MISE$.

We have assumed h to be non-random and determine it by minimizing $MISE$ given in (128). We begin with the assumption that $G_0(\cdot) = f(\cdot)$. This assumption is more pedagogical than practical but we will show subsequently that the prior

convergence rate remains unchanged under less restrictive assumptions on G_0. Of course, it is most appropriate to study the posterior convergence rate since the information contained in the data ensures that the resultant calculations are valid under far greater generality; in fact, restrictive assumptions on G_0 are not necessary at all. However, since in the Bayesian paradigm it is often recommended that the prior parameters be chosen before observing the data, it perhaps makes sense to choose values of the prior parameters which keep the prior $MISE$ rates at the desired level.

Since EW is a special case of SB, we begin by calculating the $MISE$ for the SB model; that for EW will follow quite simply then.

S-10.2.1. *Calculations.*

$$\begin{aligned} MISE &= \int_y E(\hat{f}_{SB}(y \mid \Theta_{M_n}, h) - f(y))^2 dy \\ &= \int_y \int_{\Theta_{M_n}} \left\{ \frac{1}{M_n} \sum_{l=1}^{M_n} \frac{1}{h} K\left(\frac{y-\theta_l}{h}\right) - f(y) \right\}^2 dH\left(\Theta_{M_n}\right) dy \\ &= \int_y Var\left(\hat{f}_{SB}\left(y \mid \Theta_{M_n}, h\right)\right) dy + \int_y \left\{bias_h(y)\right\}^2 dy, \end{aligned} \tag{129}$$

where $bias_h(y) = \left\{E\left(\hat{f}_{SB}\left(y \mid \Theta_{M_n}, h\right)\right) - f(y)\right\}$ and $H\left(\Theta_{M_n}\right)$ is the joint distribution of Θ_{M_n}. We write,

$$bias_h(y) = \int h^{-1} K\left(\frac{y-x}{h}\right) G_0(x) dx - f(y)$$

Calculations similar to those in Silverman (1986) yield

$$\int \left(bias_h(y)\right)^2 dy = O(h^4) \tag{130}$$

Now, the variance term can be expressed as

$$\begin{aligned} & Var(\hat{f}_{SB}(y \mid \Theta_{M_n}, h)) \\ = \; & \frac{1}{M_n^2 h^2} \sum_{j=1}^{M_n} Var\left(K\left(\frac{y-\theta_j}{h}\right)\right) + \\ & \frac{1}{M_n^2 h^2} \sum_{j=1}^{M_n} \sum_{l=1, l\neq j}^{M_n} Cov\left(K\left(\frac{y-\theta_j}{h}\right), K\left(\frac{y-\theta_l}{h}\right)\right) \end{aligned}$$

Now, assuming that h is small and M_n is large, and expanding $f(y-th)$ as a Taylor series we obtain

$$\int \frac{1}{M_n^2h^2}\sum_{j=1}^{M_n} Var\left(K\left(\frac{y-\theta_j}{h}\right)\right)dy = O\left(\frac{1}{M_nh}\right) \tag{131}$$

For the covariance term,

$$\begin{aligned} & \frac{1}{M_n^2h^2}\sum_{j}\sum_{l\neq j} Cov\left(K\left(\frac{y-\theta_j}{h}\right), K\left(\frac{y-\theta_l}{h}\right)\right) \\ = & \frac{M_n(M_n-1)}{M_n^2h^2} Cov\left(K\left(\frac{y-\theta_1}{h}\right), K\left(\frac{y-\theta_2}{h}\right)\right) \\ = & \left(1-\frac{1}{M_n}\right)\frac{1}{h^2} Cov\left(K\left(\frac{y-\theta_1}{h}\right), K\left(\frac{y-\theta_2}{h}\right)\right) \\ \leq & \frac{1}{h^2} Cov\left(K\left(\frac{y-\theta_1}{h}\right), K\left(\frac{y-\theta_2}{h}\right)\right). \end{aligned} \tag{132}$$

Simple calculations, using Taylor's series expansion yields

$$\int_y \frac{1}{h^2} Cov\left(K\left(\frac{y-\theta_j}{h}\right), K\left(\frac{y-\theta_l}{h}\right)\right)dy = O\left(\frac{1}{\alpha+1}\right). \tag{133}$$

From (130), (131) and (133) we conclude that

$$MISE(SB) = O\left(\frac{1}{M_nh} + \frac{1}{\alpha+1} + h^4\right). \tag{134}$$

We will assume M_n and α to be functions of n and will determine h as a function of n by minimizing the above form of $MISE$. The resultant form of $MISE$ will be compared with that corresponding to the EW model.

S-10.3. *Determination of h.* Minimizing (134) with respect to h yields $h_{opt} = O\left(M_n^{-\frac{1}{5}}\right)$. Plugging h_{opt} in (134) gives the optimal order of $MISE(SB)$:

$$MISE(SB)_{opt} = O\left(\frac{1}{\alpha+1} + M_n^{-\frac{4}{5}}\right) \tag{135}$$

Choosing $\alpha = O\left(M_n^{\lambda}\right)$ we get $MISE(SB)_{opt} = O\left(M_n^{-\lambda}\right)$ if $\lambda < \frac{4}{5}$ and $O\left(M_n^{-\frac{4}{5}}\right)$ if $\lambda \geq \frac{4}{5}$.

S-10.4. *Extension to a more general situation.* In the prior-based MISE calculated so far we have assumed that $G_0(\cdot) = f(\cdot)$. Let us now assume a more general situation where $G_0(\theta) = \frac{1}{M_n} g(\theta) + \left(1 - \frac{1}{M_n}\right) f(\theta)$, where $g(\cdot)$ is a density different from the true density $f(\cdot)$. It can be easily shown that here also the order of $MISE(SB)$ and $MISE(SB)_{opt}$ remain of the form as (134) and (135) respectively.

S-10.5. *Prior convergence rates of the EW model.* The form of the predictive distribution of EW is

$$\hat{f}_{EW}(y \mid \Theta_n, h) = \frac{\alpha}{\alpha+n} A + \frac{1}{\alpha+n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{y-\theta_i}{h}\right).$$

Note that the conditional distribution of θ_i given Θ_{-in} $(= \Theta_n \setminus \theta_i)$ is given by $\frac{\alpha}{\alpha+n} G_0(\theta_i) + \frac{1}{\alpha+n} \sum_{l=1, l\neq i}^{n} \delta_{\theta_l}(\theta_i)$. Suppose $\alpha = O(n^{\omega})$. If we choose $\omega > 1$, then $\frac{\alpha}{\alpha+n} \to 1$. This will imply that the conditional distribution θ_i given Θ_{-in} is close to $G_0(\theta_i)$ for large n. This fact will defeat the basic goal of using nonparametric prior. Thus we will set $\omega < 1$ and choose α such that $\frac{n}{\alpha+n} \to 1$. Thus for large n the model will look like $\frac{1}{\alpha+n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{y-\theta_i}{h}\right)$, the form similar to our form in (127). Calculations similar to those associated with the $MISE$ of the SB model show that for the EW model the $MISE$ is given by

$$MISE(EW) = O\left(\frac{1}{(\alpha+n)h} + \frac{1}{\alpha+1} + h^4\right) \tag{136}$$

Minimizing this form of $MISE$ yields h_{opt} which is of the order $\frac{1}{(\alpha+n)^{1/5}}$. Plugging h_{opt} in $MISE(EW)$ gives $MISE(EW)_{opt} = O\left(\frac{1}{(\alpha+n)^{4/5}} + \frac{1}{\alpha+1}\right)$.

Note that the order of $MISE(EW)$ is less than that of $MISE(SB)$ unless $M_n > n$. This is to be expected since so far our calculations are with respect to the prior only, not involving the data; hence, more the number of components in the mixture model faster the rate of convergence.

S-11. Comparison between prior and posterior convergence rates. To provide a feel for the prior and the posterior rates associated with the two competing models of EW and SB, we present plots of the orders of $MISE(EW)$ and $MISE(SB)$ for both prior and posterior predictive cases, for appropriate values of the parameters involved in the rates of convergence.

We assume the following forms: $M_n = n^b$, $\alpha = n^{\omega}$, $\sigma_n^2 = \frac{c^2}{4e^{n^t}}$, $\epsilon_n^* = \frac{1}{n^r}$, where $r = 3$, and $t > 0$ (the $MISE$'s will be compared for different values of t).

By equating $\epsilon^*_n = \frac{1}{n^r}$, we can calculate the value of $\frac{\epsilon_n}{1-\epsilon_n}$ which is used get closed form expression for $\epsilon^*_{M_n}$. In the above forms we assume that $b = 0.2$, $\mu_0 = 2.0$, and $\sigma_0 = 1.0$. For different values of t and ω, figures S-1 and S-2 present and compare the prior based rates of $MISE(EW)$ and $MISE(EW)$ for different values of t and r. On the other hand, figures S-3 and S-4 compare the $MISE$ rates of EW and SB based on the posterior.

As already discussed in Section S-10.5, figures S-1 and S-2 show that the prior EW rates are faster than the prior SB rates, which is expected, since, in this illustration, $M_n < n$. On the other hand, the figures S-3 and S-4 show that the posterior SB rates are far superior to the posterior EW rates.

Supplement to "Bayesian MISE convergence rates of mixture models based on the Polya urn model: asymptotic comparisons and choice of prior parameters"
(). Section S-2 contains proofs of results associated with Section 6, Section S-3 contains proofs of the results presented in Section 7, the proofs of the results provided in Section 8 are given in Section S-4, Section S-5 contains proofs of the results associated with Section 9, the proofs of the results presented in Section 12 are provided in Section S-6, Section S-7 contains a method of splitting the integrals needed for proofs of the results associated with the "large p, small n problem" in the case of the SB model. Proofs of the results stated in Section 13 are provided in Section S-8, Section S-9 contains the proofs of the results corresponding to Section 14, Section S-10 contains results and proofs associated with the Bayesian $MISE$-based prior predictive convergence rates, and finally, comparison between prior and posterior Bayesian $MISE$-based convergence rates is provided in Section S-11.

References.

BHATTACHARYA, S. (2008). Gibbs Sampling Based Bayesian Analysis of Mixtures with Unknown Number of Components. *Sankhya. Series B* **70** 133–155.

BLACKWELL, D. and MCQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Annals of Statistics* **1** 353–355.

ESCOBAR, M. D. and WEST, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association* **90** 577–588.

FERGUSON, T. S. (1983). Bayesian Density Estimation by Mixtures of Normal Distributions. In *Recent Advances in Statistics* (H. RIZVI and J. RUSTAGI, eds.) 287–302. New York: Academic Press.

GHOSAL, S. and VAN DER VAART, A. (2007). Posterior Convergence Rates of Dirichlet Mixtures At Smooth Densities. *The Annals of Statistics* **35** 697–723.

LO, A. Y. (1984). On a Class of Bayesian Nonparametric Estimates: I. Density Estimates. *The Annals of Statistics* **12** 351–357.

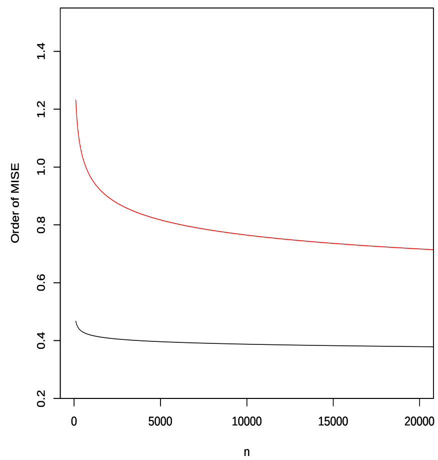

FIG S-1. *Prior predictive case with $\omega = 0.05$. The black curve corresponds to the prior predictive $MISE$ of the EW model and the red curve corresponds to that of the SB model. The left and the right panel correspond to $t = 2$ and $t = 5$, respectively.*

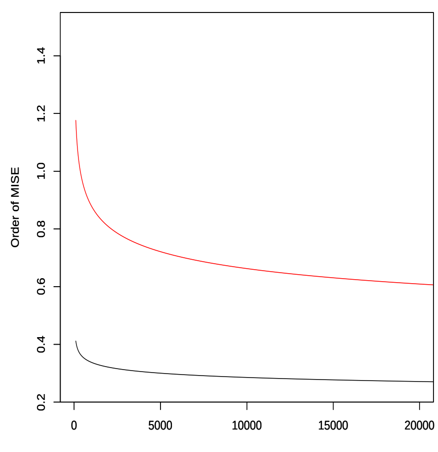

FIG S-2. *Prior predictive case with $\omega = 0.1$. The black curve corresponds to the prior predictive $MISE$ of the EW model and the red curve corresponds to that of the SB model. The left and the right panel correspond to $t = 2$ and $t = 5$, respectively.*

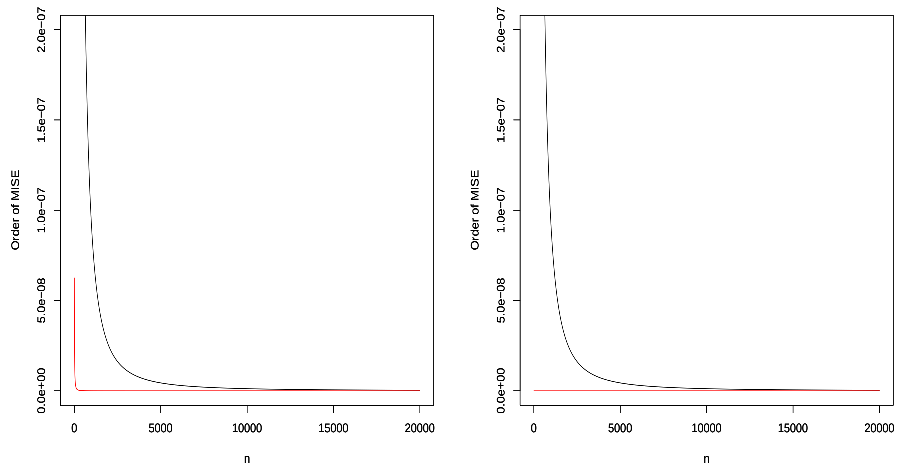

FIG S-3. *Posterior predictive case with $\omega = 0.05$. The black curve corresponds to the prior predictive $MISE$ of the EW model and the red curve corresponds to that of the SB model. The left and the right panel correspond to $t = 2$ and $t = 5$, respectively.*

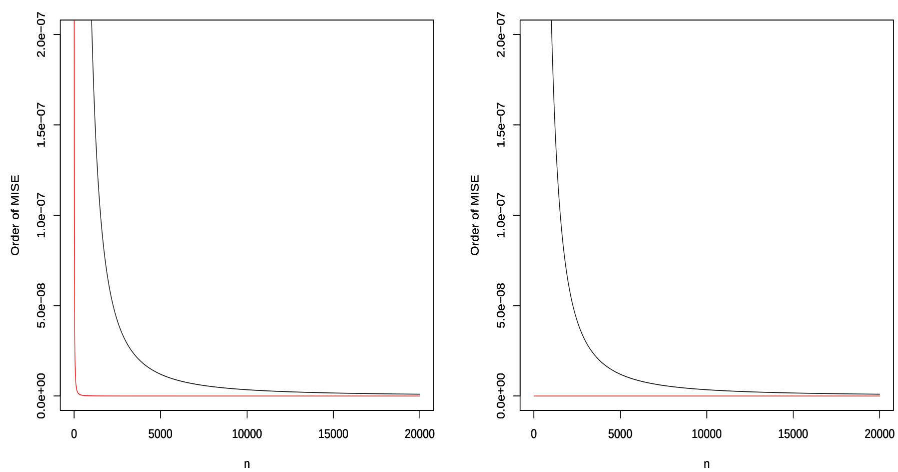

FIG S-4. *Posterior predictive case with $\omega = 0.1$. The black curve corresponds to the prior predictive $MISE$ of the EW model and the red curve corresponds to that of the SB model. The left and the right panel correspond to $t = 2$ and $t = 5$, respectively.*

MUKHOPADHYAY, S., BHATTACHARYA, S. and DIHIDAR, K. (2011). On Bayesian Central Clustering: Application to Landscape Classification of Western Ghats. *Annals of Applied Statistics* **7** 1948–1977.

MUKHOPADHYAY, S. and BHATTACHARYA, S. (2012a). Bayesian MISE Based Convergence Rates of Mixture Models Based on the Polya Urn Model: Asymptotic Comparisons and Choice of Prior Parameters. Submitted.

MUKHOPADHYAY, S. and BHATTACHARYA, S. (2012b). Supplement to "Bayesian MISE Based Convergence Rates of Mixture Models Based on the Polya Urn Model: Asymptotic Comparisons and Choice of Prior Parameters". Submitted.

MUKHOPADHYAY, S., ROY, S. and BHATTACHARYA, S. (2012). Fast and Efficient Bayesian Semi-Parametric Curve-Fitting and Clustering in Massive Data. *Sankhya. Series B*. To appear.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc., New York.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.

BAYESIAN AND INTERDISCIPLINARY RESEARCH UNIT
INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD, KOLKATA 700108
E-MAIL: sabyasachi_r@isical.ac.in

BAYESIAN AND INTERDISCIPLINARY RESEARCH UNIT
INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD, KOLKATA 700108
E-MAIL: sourabh@isical.ac.in